

Analyse des Quantisierungsfehlers bei der Verteilten Arithmetik

von
Karl-Dirk Kammeyer

Erlangen, 1977

Analyse des Quantisierungsfehlers bei der Verteilten Arithmetik

Dem Fachbereich Ingenieurwissenschaften
der Universität Erlangen - Nürnberg

zur Erlangung des Grades
Doktor - Ingenieur

vorgelegt von
Karl-Dirk Kammeyer

Erlangen, 1977



Als Dissertation genehmigt vom Fachbereich
Ingenieurwissenschaften der Universität
Erlangen - Nürnberg

Tag der Einreichung: 27. Mai 1977

Tag der Promotion: 28. Juli 1977

Dekan: Prof. Dr. rer. nat. H. J. Schneider

Berichterstatter: Prof. Dr.-Ing. W. Schüßler
Prof. Dr. rer.nat. F. Hofmann

Herrn Professor Dr.-Ing. W. Schüßler danke ich für die interessante Aufgabenstellung, für die in vielen Diskussionen gegebenen Anregungen und Hinweise und nicht zuletzt für die wertvolle Hilfe bei der Abfassung der Niederschrift.

Herrn Professor Dr. rer. nat. F. Hofmann gebührt mein Dank für sein großes Interesse an dieser Arbeit.



Digitized by the Internet Archive
in 2026

INHALTSVERZEICHNIS

	Seite
1. Einführung	5
1.1 Historischer Überblick	5
1.2 Ziel der vorliegenden Arbeit	6
2. Das Prinzip der Verteilten Arithmetik	8
2.1 Zweierkomplement-Darstellung des Eingangssignals	8
2.2 Modifizierte Zweierkomplement-Darstellung des Eingangssignals	10
3. Anwendungsbeispiele der Verteilten Arithmetik	13
3.1 Realisierung rekursiver Filter	13
3.2 Realisierung nichtrekursiver Filter	16
3.3 Allgemeine Bedeutung der Verteilten Arithmetik	20
4. Fehleranalyse für gleichverteiltes weißes Rauschen	23
4.1 Prinzipielles zum Fehlerverhalten der Verteilten Arithmetik	23
4.1.1 Darstellung der verschiedenen Fehler- einflüsse	23
4.1.2 Vergleich der Auswirkungen endlicher , Koeffizientenspeicher-Wortlängen bei Konzentrierter und Verteilter Arithmetik	25
4.2 Einfaches Rauschmodell	28
4.2.1 Fehlerleistung nach Peled und Liu	28
4.2.2 Simulation einiger Beispiele	30
4.2.3 Berechnung der exakten Fehlervarianzen für individuelle Systeme	34

4.3 Allgemeine Formulierung des Ausgangsfehlers infolge endlicher Tafel-Wortlänge	36
4.3.1 Zweierkomplement-Codierung der Eingangssignale	36
4.3.2 Modifizierte Zweierkomplement- Codierung der Eingangssignale	42
4.4 Gesamtleistung des Fehlers	44
4.4.1 Statistische Eigenschaften der Ein- gangssignale	44
4.4.2 Berechnung der zeitlichen Mittelwerte	45
4.4.3 Fehlerleistung bei Zweierkomplement-Dar- stellung der Eingangssignale	46
4.4.4 Fehlerleistung beim modifizierten Zweier- komplement	48
4.5 Definition der Rest-Koeffizientenfehler	50
4.6 Analytische Beschreibung der Rest-Koeffizienten- fehler	52
4.7 Untersuchung einzelner Beispiele	54
4.7.1 Überprüfung der Fehleranalyse durch Simulation	54
4.7.2 Einige Eigenschaften im Fehlerverhalten der Verteilten Arithmetik	59
5. Ensemble-Untersuchungen	65
5.1 Zweck der Ensemble-Untersuchungen	65
5.2 Analytische Formulierung für $r = 2$	66
5.2.1 Zweierkomplement	66
5.2.2 Modifiziertes Zweierkomplement	69
5.3 Untersuchung der Fehlerleistung	71
5.3.1 Erwartungswerte	71
5.3.2 Varianzen	74
5.3.3 Verteilungen	76

5.4 Untersuchung der Rest-Koeffizientenfehler	81
5.4.1 Varianzen	81
5.4.2 Verteilungen	84
6. Fehleranalyse bei unkorrelierten Variablen mit beliebigen statistischen Eigenschaften	87
6.1 Simulation einiger Beispiele	87
6.2 Vorbetrachtungen zur Fehleranalyse	91
6.3 Einführung der Bit-Korrelations-Matrix	93
6.4 Berechnung der Gesamtfehler-Leistung	97
6.5 Rest-Koeffizientenfehler	99
6.6 Statistisch unterschiedliche Eingangssignale	100
6.7 Analyse einzelner Beispiele	101
6.8 Ensemble-Untersuchungen	102
6.9 Abschließende Bemerkungen	107
7. Fehleruntersuchungen bei Berücksichtigung von Korrelationen zwischen den Variablen	109
7.1 Variablen-Korrelationen bei digitalen Filtern	109
7.2 Überlegungen zur Fehleranalyse bei korrelierten Variablen	112
7.2.1 Gesamte Fehlerleistung	112
7.2.2 Rest-Koeffizientenfehler	115
7.2.3 Unkorrelierte Fehlerleistung	116
7.3 Untersuchung einzelner Beispiele	116
7.3.1 Simulation zur Formulierung der Variablen-Statistik	116
7.3.2 Beispiele	120

7.4 Ensemble-Untersuchungen	126
7.4.1 Erwartungswerte der Fehlerleistungen	126
7.4.2 Rest-Koeffizientenfehler	130
7.5 Abschließende Bemerkungen	136
8. Anwendung der Fehleranalyse auf nichtrekursive Filter	138
8.1 Allgemeine Vorbetrachtungen	138
8.2 Abhängigkeit des Signal-Störverhältnisses von der Quantisierungsstufe des Koeffizientenspeichers	140
8.3 Verzerrungen des Dämpfungsverlaufs nichtrekursiver Filter infolge endlicher Koeffizientenspeicher-Wortlänge	150
8.3.1 Definition des Frequenzganges	150
8.3.2 Berechnung der Dämpfung durch Simulation	152
8.3.3 Analytische Behandlung mit dem linearen Ersatzmodell	156
8.3.4 Vergleich von Analyse und Simulation	159
8.4 Beurteilung der Realisierung nichtrekursiver Filter in Verteilter Arithmetik	163
9. Zusammenfassung	167
10. Anhang	
1. Zur Berechnung des Fehlers Δy_0	170
2. Zur Berechnung des Fehlers Δy_1	171
3. Berechnung des Ausdrucks $\langle B(n, \mathbf{x}(k)) \cdot B(m, \mathbf{x}(k)) \rangle$	172
4. Berechnung des Ausdrucks $\langle x_p(k) \cdot B(n, \mathbf{x}(k)) \rangle$	176

	Seite
5. Berechnung der Leistung N_{ϵ}	179
6. Berechnung der Verteilung von $\alpha_{n\ell}$	180
7. Analytische Überlegungen zum Erwartungswert der Fehlerleistung bei der modifizierten Form	183
8. Berechnung von $\langle x_p(k) \cdot B(n, \mathbf{x}(k)) \rangle$ bei nicht gleichverteilten Variablen	186
9. Berechnung von $\langle B(n, \mathbf{x}(k)) \cdot B(m, \mathbf{x}(k)) \rangle$ bei nicht gleichverteilten Variablen	189
10. Tabellarische Zusammenstellung der verwendeten Beispiele	191
Literaturverzeichnis	195

VERZEICHNIS DER WICHTIGSTEN FORMELZEICHEN

$a(\Omega)$	Dämpfung
$\mathbf{a} = (a_1, \dots, a_p, \dots, a_r)^T$	Koeffizientenvektor
$\mathbf{a} = (\alpha_1, \dots, \alpha_p, \dots, \alpha_r)^T$	Vektor der a-priori-Koeffizientenfehler
$\mathbf{a}^* = (\alpha_1^*, \dots, \alpha_p^*, \dots, \alpha_r^*)^T$	Vektor der Rest-Koeffizientenfehler
$B(n, \mathbf{x}(k))$	variablenabhängige Gewichte der Fehlervektor-Komponenten
β	Parameter für Verteilungsdichtefunktionen
d_{nm}	Anzahl verschiedener Bits gleicher Stellenwertigkeit von n und m
δ_{nm}	Kronecker-Symbol
$E\{N\}$	Erwartungswert von N bei Ensemble-Untersuchungen
$\mathbf{e}^{(0,1)} = (e_0^{(0,1)}, e_1^{(0,1)}, \dots, e_n^{(0,1)}, \dots, e_{2^r-1}^{(0,1)})^T$	Fehlervektor
	(Index 0 für Zweierkomplement, Index 1 für modifizierte Form)
$\epsilon_{0,1}(k)$	linear nicht korrelierter Fehleranteil (Index 0 für Zweierkomplement, Index 1 für modifizierte Form)
Φ	Bit-Korrelations-Matrix
ϕ_{ij}	Elemente der Bit-Korrelations-Matrix
$\varphi_{xx}(\kappa)$	Autokorrelationsfunktion der Variablen x(k)
$\varphi_{xy}(\kappa)$	Kreuzkorrelationsfunktion der Variablen x(k) und y(k)

φ_{∞}	Polwinkel des rekursiven Filters zur Spektralformung
$H(z)$	Übertragungsfunktion digitaler Filter
$H(\Omega)$	Frequenzgang digitaler Filter
$h(k)$	Impulsantwort nichtrekursiver Filter
i, j	Indizes verschiedener Bits einer Variablen
K	Anzahl der benutzten Abtastwerte bei der Simulation zeitlich ablaufender Vorgänge
k	normierte Zeitvariable
κ	Parameter für Korrelationsfunktionen
L	Benutzte Stichprobenmenge bei Ensemble-Untersuchungen
l	Index für die einzelnen Koeffizientensätze bei Ensemble-Untersuchungen
m, n	Indizes für Fehlervektor-Komponenten
μ_{ρ}, ν_{ρ}	jeweils ρ .Bit bei Binärdarstellung von m bzw. n
$N_{\Delta Y}, N_{\epsilon}$	Leistungen der Fehler Δy bzw. ϵ
$n_{\Delta Y}$	auf Q_R^2 bezogene Fehlerleistung
N'	Stör-Signalabstand
Ω	normierte Frequenz
$P(A)$	Wahrscheinlichkeit des Ereignisses A
$p_{\alpha}(\beta)$	Verteilungsdichtefunktion von α
Q_E, Q_A, Q_C, Q_R	Quantisierungsstufen (Eingangssignal, Ausgangs-

	signal, Koeffizientenspeicher bei Konzentrierter und Verteilter Arithmetik)
R	gesamte Längen der zu multiplizierenden Vektoren (Filterlänge bei nichtrekursiven Filtern)
R_E	Rauschzahl
r_λ, r	Länge der Teilsummen bei Anwendung der Verteilten Arithmetik
ρ_∞	Polradius des rekursiven Filters zur Spektralformung
ρ, σ	Indizes der Koeffizienten- bzw. Variablenvektor-Komponenten
$\left. \begin{matrix} \sigma_x^2, \sigma_{\Delta u}^2 \\ \sigma_{\Delta v}^2, \sigma_{\Delta y}^2 \end{matrix} \right\}$	Varianzen der zeitabhängigen Größen $x, \Delta u, \Delta v, \Delta y$
$u(\xi_{1i}, \dots, \xi_{ri})$	Koeffizientenspeicherinhalt bei Verteilter Arithmetik, wenn Variablen in gewöhnlichem Zweierkomplement codiert werden
Δu_i	Quantisierungsfehler der Größe $u(\xi_{1i}, \dots, \xi_{ri})$
$v(\xi_{1i}, \dots, \xi_{ri})$	Koeffizientenspeicherinhalt bei Verteilter Arithmetik, wenn Variablen im modifizierten Zweierkomplement codiert werden
Δv_i	Quantisierungsfehler der Größe $v(\xi_{1i}, \dots, \xi_{ri})$
$\text{VAR}\{N\}$	Varianz von N bei Ensemble-Untersuchungen
w, w_c, w_R	Wortlängen der Variablen, Koeffizientenspeicher bei Konzentrierter und Verteilter Arithmetik
Δw	erforderliche innere Wortlängen-Erhöhung
$\mathbf{x} = (x_1, x_2, \dots, x_\rho, \dots, x_r)$	Variablenvektor
Δx_ρ	Quantisierungsfehler der Variablen x_ρ
x_{f1}, x_{f2}, x_{f3}	Drei verschiedene Signaltypen gefärbten Rauschens

x_w	weiße Rauschgröße
$\xi_{\rho i}$	i.Bit der Variablen $x_{\rho}(k)$
y	Ergebnis des inneren Produktes zweier Vektoren bei unverkürzten Wortlängen
$\Delta y_E, \Delta y_A, \Delta y_C$	Verschiedene Fehlereinflüsse infolge Ein- und Ausgangsquantisierung und endlicher Koeffizienten- wortlänge bei Konzentrierter Arithmetik
$\Delta y_0, \Delta y_1$	Fehler von y infolge endlicher Wortlänge des Ko- effizientenspeichers bei Verteilter Arithmetik (Index 0 bei Zweierkomplement, Index 1 bei modi- fizierter Form)
Δy_{res}	Resultierender Fehler bei Berücksichtigung aller Fehlereinflüsse
$\mathbb{Z}$	Menge der ganzen Zahlen
z	Komplexe Variable der z-Transformation

Sonderzeichen

$[a]$	größte ganze Zahl, die kleiner oder gleich a ist
$[a]_Q$	Auf die Quantisierungsstufe Q gerundete Größe a
$\langle x(k) \rangle$	zeitlicher Mittelwert der Größe $x(k)$
$\bar{\xi}$	Inversion der Binärvariablen ξ
$\xi_1 \oplus \xi_2$	Antivalenz-Verknüpfung der Binärvariablen ξ_1 und ξ_2
$\xi_1 \odot \xi_2$	Valenz-Verknüpfung der Binärvariablen ξ_1 und ξ_2

1. EINFÜHRUNG

1.1 Historischer Überblick

Zur Realisierung digitaler Filter wurde 1971 von E. Andersson [1] ein neues Verfahren angegeben, bei dem zur Berechnung der Summe von Produkten keine Multiplizierer im herkömmlichen Sinne mehr benutzt werden. Statt dessen wird die Struktur einer arithmetischen Einheit angegeben, die jeweils Bit-Gruppen gleicher Binärstellenwertigkeit von allen Eingangsvariablen zusammenfaßt und daraus zugehörige Koeffizientenkombinationen berechnet. Das Endergebnis wird durch bitweise verschobene Akkumulation gewonnen.

Damit war zum ersten Mal ein Prinzip formuliert, dessen Bedeutung weit über die in der zitierten Arbeit beschriebene spezielle Filterrealisierung hinausgeht. Bei Erscheinen der Arbeit wurde das neue Verfahren kaum beachtet, bis es einige Zeit danach von mehreren Autoren unabhängig voneinander wiederentdeckt [2,3,4,5] und dabei seine Bedeutung für viele Realisierungsprobleme der Signalverarbeitung erkannt wurde. Da inzwischen eine außerordentliche Entwicklung der Halbleiter-Technologie, insbesondere der hochintegrierten Speicher, stattgefunden hatte, wurden in diesen Arbeiten zur Bildung von Koeffizienten-Partialsummen nicht mehr wie bei Andersson Addierer-Netzwerke vorgeschlagen, sondern statt dessen ROMs (read-only-memories) oder RAMs (random-access-memories) eingesetzt, die entsprechend einer Funktionstafel von den zu verarbeitenden Bit-Gruppen der Eingangssignale adressiert werden.

Von den vier zwischen 1973 und 1974 erschienenen Arbeiten ist die von Peled und Liu am bekanntesten geworden, so daß das Verfahren vielfach mit "Peled-Liu-Struktur" bezeichnet wurde. Die Bezeichnung "Verteilte Arithmetik" wurde erstmals von Schüßler [6] formuliert. Sie soll die Arbeitsweise charakterisieren, bei der nicht mehr jeder einzelnen arithmetischen Operation ein bestimmtes Element wie Multiplizierer oder Addierer zugeordnet werden kann, sondern die Abarbeitung bitweise auf alle Variablen gleichmäßig verteilt durchgeführt wird. Die herkömmliche

Technik wird im Gegensatz dazu mit "Konzentrierte Arithmetik" bezeichnet.

1.2 Ziel der vorliegenden Arbeit

Inzwischen sind in einer Reihe von weiteren Veröffentlichungen Anwendungsmöglichkeiten und Realisierungsprobleme der Verteilten Arithmetik ausführlich erörtert worden [z.B. 7,8,9,10,11, 12]. Dabei erwies sich dieses Verfahren in allen Fällen als sehr attraktiv, was die Aspekte einer günstigen Schaltungsrealisierung betrifft. Ein anderes wesentliches Beurteilungskriterium, die Auswirkungen von Quantisierungsfehlern, wurde bisher nur in Form eines stark vereinfachten Rauschmodells behandelt [5]. Das Ziel der vorliegenden Arbeit ist daher eine eingehende Analyse des Quantisierungsfehlers bei der Verteilten Arithmetik.

In Kapitel 2 wird zunächst das Prinzip der "Verteilten Arithmetik" erläutert. Daran schließen sich in Kapitel 3 einige Anwendungsbeispiele an, wobei auf Realisierungsdetails nur soweit eingegangen wird, wie sie für das Fehlerverhalten von Bedeutung sind.

In Kapitel 4 wird für gleichverteiltes weißes Eingangsrauschen eine exakte Analyse des Quantisierungsfehlers entwickelt. Dabei interessieren neben der gesamten Fehlerleistung die Koeffizientenfehler, die für den Fall der Verteilten Arithmetik neu definiert werden müssen und mit "Rest-Koeffizientenfehler" bezeichnet werden. Es ergibt sich aus diesen Untersuchungen eine Aufspaltung in linear mit den Eingangssignalen korrelierte Anteile und eine unkorrelierte Rauschgröße.

Um über das Fehlerverhalten der Verteilten Arithmetik prinzipielle Aussagen machen zu können, werden in Kapitel 5 große Ensembles von Systemen mit zufällig gewählten Koeffizientensätzen betrachtet. Daraus sind statistische Kenngrößen der Fehlerparameter wie Erwartungswerte der Fehlerleistung, Verteilungen der Restkoeffizientenfehler usw. zu ermitteln.

Es wird sich im weiteren erweisen, daß das Fehlerverhalten der Verteilten Arithmetik von den statistischen Eigenschaften der Eingangssignale abhängt. Die in Kapitel 4 und 5 gefundenen Ergebnisse werden deshalb, soweit möglich, auf allgemeinere Signale erweitert. So enthält Kapitel 6 eine exakte Analyse für beliebig verteilte, aber unkorrelierte Eingangssignale und anschließend die oben für gleichverteiltes Rauschen erläuterten Ensemble-Untersuchungen.

In Kapitel 7 schließlich wird der Fall beliebig korrelierter Signale betrachtet. Die Formulierung eines geschlossenen analytischen Ausdruckes gelingt hier nicht mehr, aber immerhin ermöglicht die im Ansatz enthaltene Trennung von Signal- und Systemparameter-Einflüssen auf das Fehlersignal die Untersuchung von großen Ensembles, ohne daß man auf die Simulation der einzelnen Systeme angewiesen ist.

In Kapitel 8 schließlich soll versucht werden, die abgeleiteten Formeln und entwickelten Verfahren der Fehleranalyse bei der Untersuchung nichtrekursiver Filter anzuwenden. Von besonderem Interesse sind dabei die linearen Verzerrungen des Frequenzganges. Die analytisch berechenbare Frequenzgang-Verformung mit Hilfe der oben genannten Rest-Koeffizientenfehler wird dem durch Simulation bestimmten Frequenzgang gegenübergestellt.

2. DAS PRINZIP DER VERTEILTEN ARITHMETIK

2.1 Zweierkomplement-Darstellung des Eingangssignals

Von White [13] wurde erstmals darauf hingewiesen, daß die Verteilte Arithmetik allgemein als Realisierungsform des inneren Produktes zweier Vektoren betrachtet werden kann. Dabei ist die Realisierung digitaler Filter ein Spezialfall.

Es seien zwei Vektoren

$$\mathbf{a} = (a_1, a_2, \dots, a_\rho, \dots, a_r)^T \quad (2.1a)$$

und $\mathbf{x} = (x_1, x_2, \dots, x_\rho, \dots, x_r)^T$

vorgegeben, die skalar miteinander multipliziert werden sollen.

$$y = \mathbf{a}^T \mathbf{x} = \sum_{\rho=1}^r a_\rho x_\rho \quad (2.2)$$

Die Werte x_ρ sollen in Zweierkomplement-Darstellung mit einer Wortlänge w vorliegen und auf den Bereich $-1 \leq x_\rho < 1$ beschränkt sein.

$$x_\rho = -\xi_{\rho 0} + \sum_{i=1}^{w-1} \xi_{\rho i} 2^{-i} \quad ; \quad \xi_{\rho i} \in \{0, 1\} \quad (2.3)$$

Die Größe $\xi_{\rho i}$ in Gl. (2.3) bezeichnet das Bit der Binärstelle i der Variablen x_ρ , beginnend mit $\xi_{\rho 0}$ als Vorzeichenbit.

Gleichung (2.3) wird in (2.2) eingesetzt:

$$y = \sum_{\rho=1}^r a_\rho \left(-\xi_{\rho 0} + \sum_{i=1}^{w-1} \xi_{\rho i} 2^{-i} \right) \quad (2.4)$$

Nach Vertauschung der Summations-Reihenfolge erhält man eine Formulierung für das innere Produkt (2.2), die auf das Verfahren der Verteilten Arithmetik führt.

$$y = - \left(\sum_{\rho=1}^r a_\rho \cdot \xi_{\rho 0} \right) + \sum_{i=1}^{w-1} 2^{-i} \left(\sum_{\rho=1}^r a_\rho \cdot \xi_{\rho i} \right) \quad (2.5)$$

Dabei ist entscheidend, daß die in Klammern stehenden Ausdrücke jeweils von r binären Variablen abhängen und daher nur maximal 2^r verschiedene Werte annehmen können. Man definiert eine Funktion

$$u(\xi_{1i}, \dots, \xi_{ri}) = \sum_{\rho=1}^r a_{\rho} \xi_{\rho i} \quad (2.6)$$

mit der das innere Produkt neu ausgedrückt werden kann:

$$y = -u(\xi_{10}, \dots, \xi_{r0}) + \sum_{i=1}^{w-1} 2^{-i} u(\xi_{1i}, \dots, \xi_{ri}) \quad (2.7)$$

Gleichung (2.7) stellt die grundsätzliche Formulierung der Verteilten Arithmetik dar. Bei einer Realisierung kann danach die Funktion $u(\xi_{1i}, \dots, \xi_{ri})$ in einen Speicher eingeschrieben werden, der durch die r Binärvariablen $\xi_{1i}, \dots, \xi_{ri}$ adressiert wird. Das Ergebnis y erhält man gemäß Gl. (2.7) dann, indem die Teilergebnisse $u(\xi_{1i}, \dots, \xi_{ri})$ für alle Stellenwertigkeiten $i = 0, \dots, w-1$ gebildet und um jeweils 1 Bit versetzt aufakkumuliert werden. Eine Sonderstellung nimmt hier der Schritt $i = 0$ ein, für den das Vorzeichen von $u(\xi_{10}, \dots, \xi_{r0})$ invertiert werden muß. Die Realisierung des inneren Produktes in Verteilter Arithmetik erfordert also im wesentlichen einen Speicher und einen Akkumulator.

Betrachtet man die Definitionsgleichung für die Funktion u (Gl. 2.6), so zeigt sich, daß der Inhalt des Speichers aus sämtlichen möglichen, additiv verknüpften Koeffizientenkombinationen besteht. Zur Verdeutlichung ist in Tabelle 1.1 der Speicherinhalt für das Beispiel $r = 4$ in Abhängigkeit der 4 Adressen-Bits dargestellt.

ξ_1	ξ_2	ξ_3	ξ_4	$u(\xi_1, \xi_2, \xi_3, \xi_4)$	ξ_1	ξ_2	ξ_3	ξ_4	$u(\xi_1, \xi_2, \xi_3, \xi_4)$
0	0	0	0	0	0	0	0	1	a_4
1	0	0	0	a_1	1	0	0	1	a_1+a_4
0	1	0	0	a_2	0	1	0	1	a_2+a_4
1	1	0	0	a_1+a_2	1	1	0	1	$a_1+a_2+a_4$
0	0	1	0	a_3	0	0	1	1	a_3+a_4
1	0	1	0	a_1+a_3	1	0	1	1	$a_1+a_3+a_4$
0	1	1	0	a_2+a_3	0	1	1	1	$a_2+a_3+a_4$
1	1	1	0	$a_1+a_2+a_3$	1	1	1	1	$a_1+a_2+a_3+a_4$

Tabelle 1.1: Koeffizientenspeicher-Inhalt für $r = 4$

2.2 Modifizierte Zweierkomplement-Darstellung des Eingangssignals

In der schon zitierten Arbeit von Croisier et al. [2] wurde eine interessante Variante der Eingangssignal-Codierung eingeführt, die für die Realisierung kaum einen weiteren Aufwand bedeutet, aber den Speicherplatz auf die Hälfte reduziert. Dazu geht man von der Identität (2.8) aus,

$$x = \frac{1}{2} (x - (-x)) \quad (2.8)$$

die nach Einsetzen der Zweierkomplement-Codierung auf eine neue Darstellung der Zahl x führt:

$$x = \frac{1}{2} (-\xi_0 + \sum_{i=1}^{w-1} \xi_i 2^{-i} - (-\bar{\xi}_0 + \sum_{i=1}^{w-1} \bar{\xi}_i 2^{-i} + 2^{-w+1})) \quad (2.9a)$$

$$x = -\frac{1}{2} (\xi_0 - \bar{\xi}_0) + \frac{1}{2} \sum_{i=1}^{w-1} (\xi_i - \bar{\xi}_i) 2^{-i} - 2^{-w} \quad (2.9b)$$

mit $\bar{\xi}_i$ als Inversion der Binärvariablen ξ_i .

In Gl. (2.9b) wird die Größe x durch die Binärvariablen $(\xi_i - \bar{\xi}_i) \in \{-1, 1\}$ beschrieben.

Diese Darstellung wird auf Gl. (2.2) angewendet.

$$y = -\frac{1}{2} \sum_{\rho=1}^r (\xi_{\rho 0} - \bar{\xi}_{\rho 0}) a_{\rho} + \frac{1}{2} \sum_{i=1}^{w-1} 2^{-i} \cdot \sum_{\rho=1}^r (\xi_{\rho i} - \bar{\xi}_{\rho i}) a_{\rho} - 2^{-w} \sum_{\rho=1}^r a_{\rho} \quad (2.10)$$

In Abhängigkeit der r Bits mit 2^r möglichen Kombinationen wird wieder eine Funktion eingeführt:

$$v(\xi_{1i}, \dots, \xi_{ri}) = \sum_{\rho=1}^r (\xi_{\rho i} - \bar{\xi}_{\rho i}) a_{\rho}, \quad (2.11)$$

die in Gl. (2.10) eingesetzt wird.

$$y = -\frac{1}{2} v(\xi_{10}, \dots, \xi_{r0}) + \frac{1}{2} \sum_{i=1}^{w-1} 2^{-i} \cdot v(\xi_{1i}, \dots, \xi_{ri}) - 2^{-w} v(1, 1, \dots, 1) \quad (2.12)$$

Wegen des endlichen Wertevorrats kann $v(\xi_{1i}, \dots, \xi_{ri})$ in einen Speicher eingeschrieben werden; Tabelle 1.2 zeigt für das Beispiel $r = 4$ den Inhalt dieses Speichers.

Es zeigt sich, daß hier im Gegensatz zur gewöhnlichen Zweierkomplement-Codierung bei jeder Kombination alle Koeffizienten auftreten, die Variation besteht nur in den Vorzeichen. Bemerkenswert ist nun die Symmetrie der abzuspeichernden Werte, die durch die Anordnung der Tabelle 1.2 deutlich wird: Die linke Hälfte enthält die gleichen Werte mit negativem Vorzeichen wie die rechte. Da bei einer Realisierung die Umkehrung des Vorzeichens keinen nennenswerten Aufwand bedeutet, kann man sich auf die Abspeicherung des halben Wertevorrats beschränken und so den Gesamtaufwand erheblich reduzieren.

ξ_1 ξ_2 ξ_3 ξ_4	$v(\xi_1, \xi_2, \xi_3, \xi_4)$	ξ_1 ξ_2 ξ_3 ξ_4	$v(\xi_1, \xi_2, \xi_3, \xi_4)$
0 0 0 0	$-(a_1+a_2+a_3+a_4)$	1 1 1 1	$(a_1+a_2+a_3+a_4)$
1 0 0 0	$-(-a_1+a_2+a_3+a_4)$	0 1 1 1	$(-a_1+a_2+a_3+a_4)$
0 1 0 0	$-(a_1-a_2+a_3+a_4)$	1 0 1 1	$(a_1-a_2+a_3+a_4)$
1 1 0 0	$-(-a_1-a_2+a_3+a_4)$	0 0 1 1	$(-a_1-a_2+a_3+a_4)$
0 0 1 0	$-(a_1+a_2-a_3+a_4)$	1 1 0 1	$(a_1+a_2-a_3+a_4)$
1 0 1 0	$-(-a_1+a_2-a_3+a_4)$	0 1 0 1	$(-a_1+a_2-a_3+a_4)$
0 1 1 0	$-(a_1-a_2-a_3+a_4)$	1 0 0 1	$(a_1-a_2-a_3+a_4)$
1 1 1 0	$-(-a_1-a_2-a_3+a_4)$	0 0 0 1	$(-a_1-a_2-a_3+a_4)$

Tabelle 1.2: Koeffizientenspeicher-Inhalt bei modifiziertem Zweierkomplement ($r = 4$)

Unter Verwendung der Antivalenz-Verknüpfung zweier Binärvariablen

$$\xi_1 \oplus \xi_2 = \xi_1 \bar{\xi}_2 \vee \bar{\xi}_1 \xi_2 \quad (2.13)$$

läßt sich das innere Produkt nach Gl. (2.2) für die modifizierte Zweierkomplement-Codierung allgemein formulieren:

$$\begin{aligned}
 y = & + \frac{1}{2} (\xi_{ro} \bar{\xi}_{ro}) \cdot v((\xi_{10} \oplus \xi_{ro}), \dots, (\xi_{r-1,0} \oplus \xi_{ro}), 0) \\
 & - \frac{1}{2} \sum_{i=1}^{w-1} 2^{-i} (\xi_{ri} \bar{\xi}_{ri}) \cdot v((\xi_{1i} \oplus \xi_{ri}), \dots, (\xi_{r-1,i} \oplus \xi_{ri}), 0) \\
 & + \frac{1}{2} \cdot 2^{-w+1} v(0, 0, \dots, 0)
 \end{aligned} \quad (2.14)$$

3. ANWENDUNGSBEISPIELE DER VERTEILTEN ARITHMETIK

3.1 Realisierung rekursiver Filter

In den ursprünglichen Arbeiten [1,...,5] wurde die Verteilte Arithmetik ausschließlich als eine neue Realisierungsform rekursiver Filter verstanden.

Die Differenzengleichung eines Blockes 2. Grades

$$y(k) = b_2x(k) + b_1x(k-1) + b_0x(k-2) - c_1y(k-1) - c_0y(k-2) \quad (3.1)$$

ist direkt in Verteilter Arithmetik zu lösen, wenn die Variablen $x(k)$, $x(k-1)$, $x(k-2)$, $y(k-1)$, $y(k-2)$ im Zweierkomplement codiert werden. Bild 3.1 zeigt eine solche Realisierung.

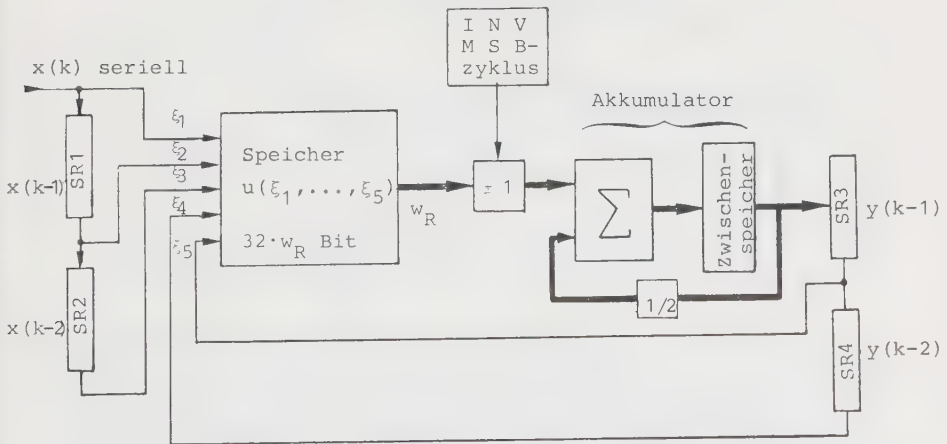


Bild 3.1: Block 2. Grades in Verteilter Arithmetik

Die Schaltung besteht aus vier Zustandsspeichern (Schieberegister SR1...SR4), dem Akkumulator mit Parallel-Addierer, Zwischen-speicher und verdrahteter Division durch 2, sowie dem Speicher für die Koeffizientenkombinationen $u(\xi_1, \dots, \xi_5)$, der bei unver-

änderlichen Koeffizienten durch ein ROM ("read-only-memory") realisiert wird. Im Falle eines flexiblen Filters kann das ROM durch ein RAM ("random-access-memory") ersetzt werden. Zur Veränderung der fünf Koeffizienten müssen dann allerdings alle 32 Koeffizientenkombinationen neu eingelesen werden, was in bestimmten Fällen, z.B. im adaptiven Echtzeit-Betrieb, unerwünscht sein kann. Um dieses zu vermeiden, bietet sich die in [1] und [7] vorgeschlagene Methode an, bei der nur die fünf Koeffizienten gespeichert und ihre verschiedenen Kombinationen in einem Schaltnetz gebildet werden. Bild 3.2 zeigt dieses Schaltnetz für das Beispiel von vier Koeffizienten.

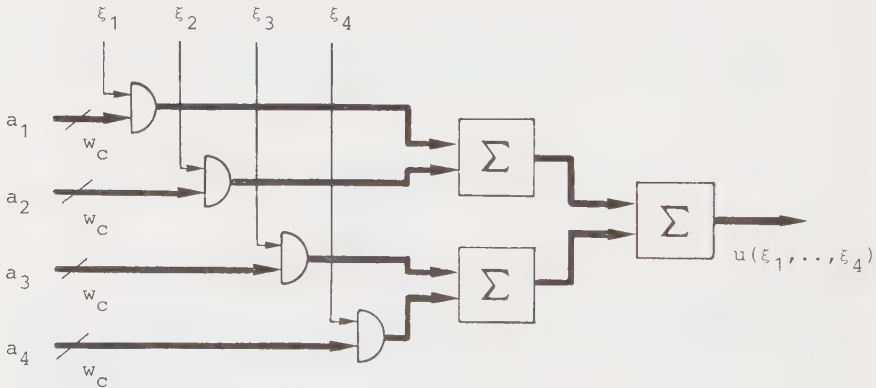


Bild 3.2: Schaltnetz zur Berechnung von $u(\xi_1, \xi_2, \xi_3, \xi_4)$

Bei Anwendung der modifizierten Zweierkomplement-Codierung ändert sich das Blockschaltbild 3.1 geringfügig: Die Variable $y(k-2)$ wird zur Vorzeichensteuerung des Speicher-Ausgangssignals benutzt, während die restlichen 4, mit $y(k-2)$ bitweise exklusiv-oder-verknüpft, die Speicheradresse bilden. Die Speicherkapazität ist damit auf $16 \cdot w_R$ Bit reduziert. Als zusätzlicher Aufwand ist dem Akkumulationszyklus ein weiterer Schritt voranzustellen, in dem der Wert $2^{-w+1}v(0, \dots, 0)$ zu laden ist (vgl. Gleichung 2.14). Schließlich ist das Endresultat durch 2 zu dividieren, was fest verdrahtet möglich ist.

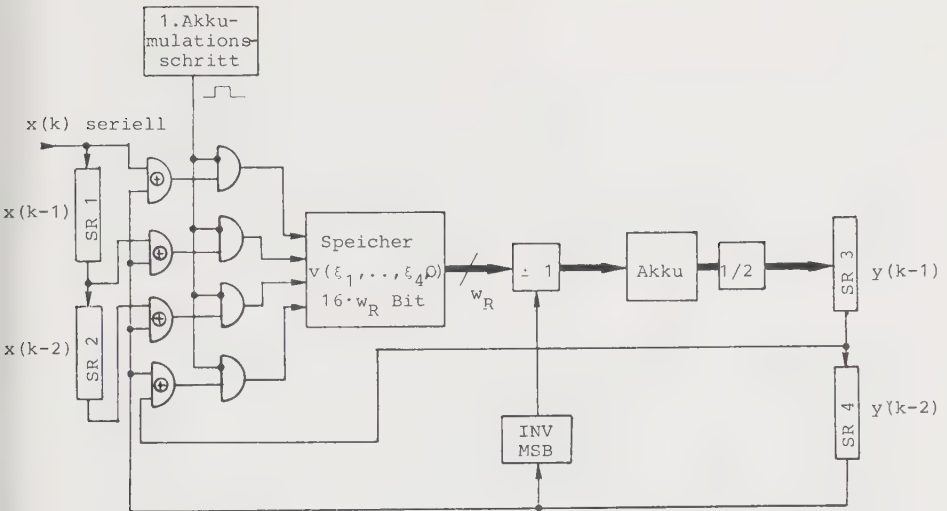


Bild 3.3: Block 2. Grades in Verteilter Arithmetik mit modifizierter Zweierkomplement-Codierung

Die bisher beschriebenen Realisierungsstrukturen des Blockes 2. Grades enthalten für die Zustandsvariablen jeweils 4 Register, sie sind also bezüglich der Zustandsspeicher nicht kanonisch. Es lassen sich jedoch nach einem Vorschlag von Claassen, Mecklenbräuker und Peek [9] auch kanonische Strukturen in Verteilter Arithmetik realisieren: Bild 3.4a,b zeigt Prinzipschaltbilder für die 1. und 2. kanonische Form.

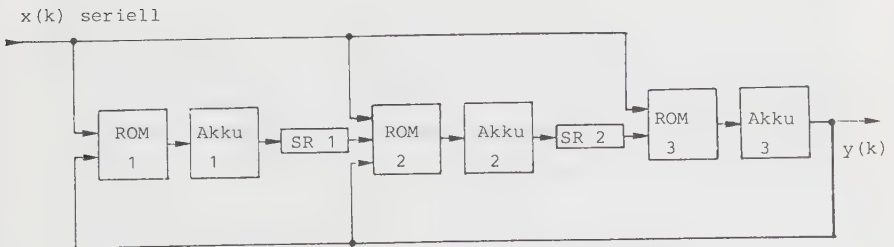


Bild 3.4a: 1. kanonische Form in Verteilter Arithmetik

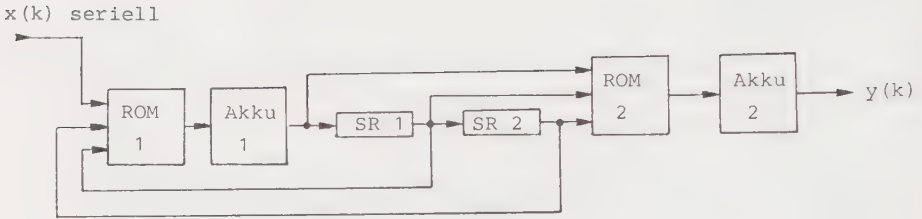


Bild 3.4b: 2. kanonische Form in Verteilter Arithmetik

Die verschiedenen Koeffizientenkombinationen sind hier nicht mehr wie bei der direkten Realisierung nach Bild 3.1 und 3.3 in einem Speicher konzentriert, sondern auf 2 bzw. 3 einzelne ROMs verteilt. Insgesamt ergibt sich hier die halbe Speicherkapazität von $16 \cdot w_R$ Bit bzw. $8 \cdot w_R$ Bit bei der modifizierten Form.

Die Einsparung von Zustands- und Koeffizientenspeicherplatz bringt den Nachteil der zwei- bzw. dreifachen Ausführung von einzelnen Akkumulatoren mit sich, was höheren Schaltungsaufwand bzw. Geschwindigkeitsverlust durch Zeitmultiplex-Verarbeitung bedeutet.

Der Vergleich verschiedener Hardware-Strukturen ist nicht der Gegenstand dieser Arbeit. Es sei hier deshalb auf eingehende Aufwandsvergleiche verschiedener Filter-Realisierungen verwiesen, die an anderer Stelle durchgeführt wurden [8,14,15]. Dabei erwies sich die Realisierung nach Bild 3.4b, also die 2. kanonische Form trotz des höheren Akkumulationsaufwandes der direkten Struktur hinsichtlich der Bauteile-Aufwendungen als überlegen.

3.2 Realisierung nichtrekursiver Filter

Die Differenzengleichung eines nichtrekursiven Filters vom Grad $R-1$ (Filterlänge R)

$$y(k) = \sum_{\rho=0}^{R-1} h(\rho) x(k-\rho) \quad (3.2)$$

ist prinzipiell in Verteilter Arithmetik zu realisieren, indem die Zustandsvariablen $x(k) \dots x(k-(R-1))$ bitweise zur Adressierung des Speichers $u(\xi_0, \dots, \xi_{R-1})$ bzw. $v(\xi_0, \dots, \xi_{R-1})$ benutzt werden.

Im allgemeinen ist der Grad $(R-1)$ so hoch, daß der erforderliche ROM- (oder RAM-) Speicherplatz von $2^R \cdot w_R$ Bit zu unrealistischen Kapazitäten führen würde.

Die Summe in Gl. (3.2) wird deshalb in L Teilsummen der Längen r_λ aufgespalten:

$$\begin{aligned} y(k) = & \sum_{\rho=0}^{r_1-1} h(\rho) x(k-\rho) + \sum_{\rho=r_1}^{r_1+r_2-1} h(\rho) x(k-\rho) + \dots \\ & + \sum_{\rho=R_\ell-1}^{R_\ell-1} h(\rho) x(k-\rho) + \dots + \sum_{\rho=R-R_L}^{R-1} h(\rho) x(k-\rho) \end{aligned} \quad (3.3)$$

mit
$$R_\ell = \sum_{\lambda=1}^{\ell} r_\lambda$$

Die Teillängen r_λ müssen dabei so gewählt werden, daß eine Realisierung durch verfügbare Speicher möglich ist. Das Endergebnis $y(k)$ erhält man dann gemäß Gl. (3.3) durch Aufaddition der Teilsummen, was je nach Geschwindigkeitsanforderungen teils durch parallele Addierer, teils durch Zeit-Multiplex-Überlagerung im Akkumulator geschehen kann.

Die folgenden Bilder zeigen die beiden Extremfälle der reinen Parallelverarbeitung (Bild 3.5a) und der Realisierung durch ein einziges ROM, bei dem zeitlich nacheinander durch einen "Multiplex-Zähler" Teilsummen im ROM angewählt werden, die der Akkumulator dann aufaddiert (Bild 3.5b). In diesem Fall arbeitet die Schaltung um den Faktor $L = R/3$ langsamer als die Parallel-Version, wenn von gleicher ROM-Auslesefrequenz ausgegangen wird, vermeidet aber den Aufwand von $L-1 = R/3 - 1$ parallelen Addierern.

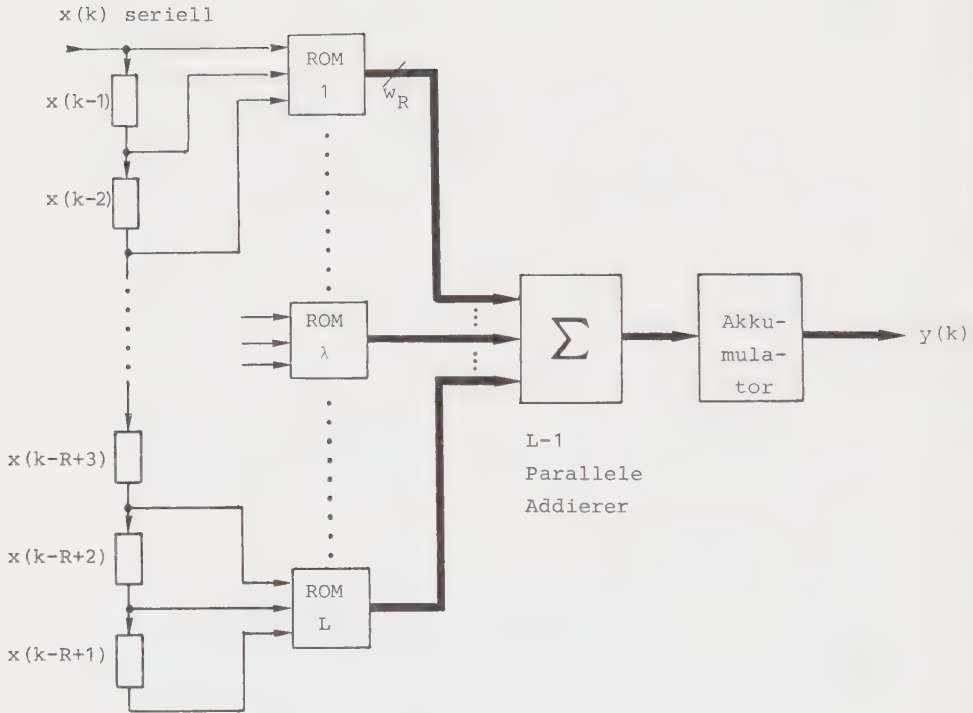


Bild 3.5a: Blockschaltbild eines nichtrekursiven Filters in Verteilter Arithmetik bei paralleler Zusammenfassung der Teilsummen
(Beispiel für $r_\lambda = 3$, $\lambda \in \{1, \dots, R/3\}$)

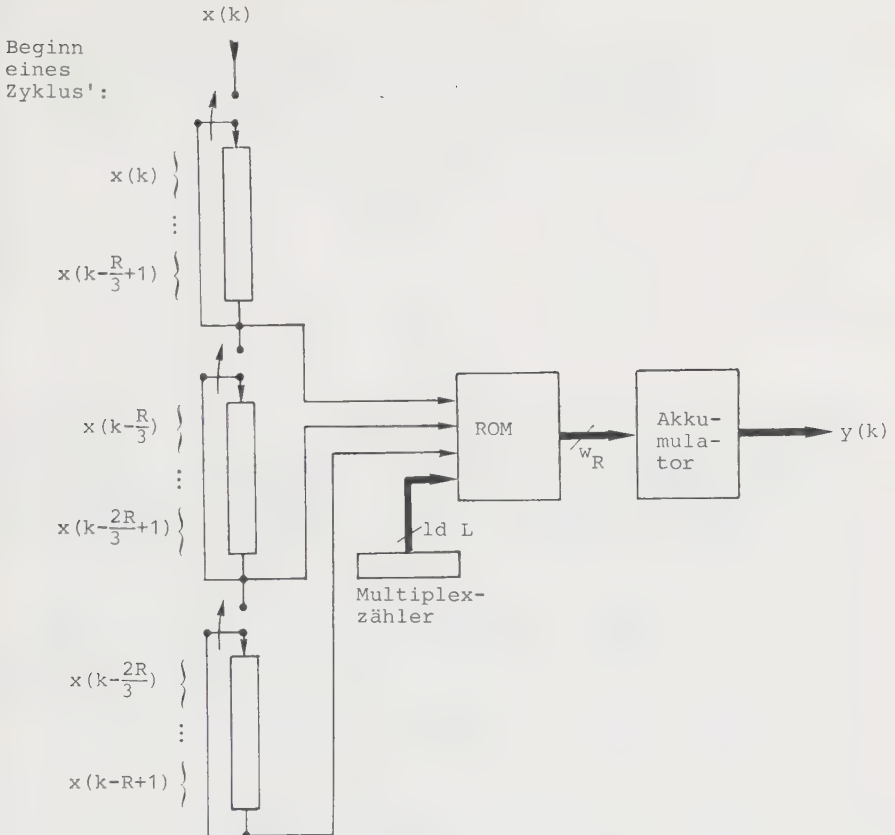


Bild 3.5b: Nichtrekursives Filter in Verteilter Arithmetik mit Multiplex-Überlagerung der Teilsummen (Beispiel $r_\lambda = 3, \lambda \in \{1, \dots, R/3\}$)

Von besonderem Interesse für das Fehlerverhalten ist hier die Organisation der Zustandsspeicher und im Zusammenhang damit die Verteilung der R Variablen auf die ROM-Eingänge. Man betrachte zunächst die Parallel-Realisierung. Hier ist die Zuordnung der Schieberegister-Ausgänge zu den ROMs frei wählbar, denn die Summanden in Gl. (3.3) sind beliebig vertauschbar.

Für die Multiplex-Struktur nach Bild 3.5b gilt dieses nicht. Die einzelnen Schieberegister des Zustandsspeichers enthalten jeweils R/r Variablen. Dabei liegt die zeitliche Reihenfolge

der an den Abgriffen erscheinenden Variablen fest. Die sich zwangsläufig ergebenden Gruppierungen am ROM-Eingang sind also im vorliegenden Beispiel

$$\begin{array}{lll} \{x(k - \frac{R}{3} + 1) & x(k - \frac{2R}{3} + 1) & x(k - R + 1)\} ; \\ \{x(k - \frac{R}{3} + 2) & x(k - \frac{2R}{3} + 2) & x(k - R + 2)\} ; \\ & \vdots & \\ \{x(k) & x(k - R/3) & x(k - \frac{2R}{3})\} . \end{array}$$

Man erkennt, daß die Variablengruppen aus Abtastwerten bestehen, die jeweils um $\frac{R}{3}$ Takte auseinanderliegen. Eine solche Gruppierung hat Auswirkungen auf die Korrelationen zwischen den gleichzeitig verarbeiteten Signalwerten. In Kapitel 7 wird gezeigt, daß diese Eigenschaften das Fehlerverhalten stark beeinflussen können.

Auf weitere Einzelheiten von günstigen Realisierungen nichtrekursiver Filter soll hier nicht eingegangen werden: Schaltungs-details, Aufwandsabschätzungen einschließlich Vergleiche mit konventionellen Strukturen wurden in [8] ausführlich erörtert.

3.3 Allgemeine Bedeutung der Verteilten Arithmetik

Es hat sich seit der ersten Formulierung des Prinzips der Verteilten Arithmetik in den fünf zitierten Arbeiten [1,...,5] gezeigt, daß diese Technik für die Signalverarbeitung von grundlegender, über die bloße Realisierung digitaler Filter hinausgehender Bedeutung ist. Mit der allgemeineren Interpretation als Lösung für das innere Produkt zweier Vektoren [13] kann sie überall dort eingesetzt werden, wo eine Summe von Produkten zu berechnen ist. Daß dieses eine oft wiederkehrende Problemstellung ist, zeigt eine Reihe von Publikationen, in denen die verschiedensten Anwendungsbeispiele der Verteilten Arithmetik vorgeschlagen werden [9,10,12,16,17,18].

Eine sehr allgemeine und theoretisch interessante Betrachtungsweise liefert eine zusammenfassende Arbeit über die Verteilte

Arithmetik von Burrus [19]. Hier wird die Faltung von zwei Zahlenfolgen zurückgeführt auf grundsätzliche Betrachtungen bis zur Ebene einzelner Bits. Anhand dieser Überlegungen erkennt man drei prinzipiell mögliche Stufen von Verarbeitungs-Ebenen, denen jeweils bekannte Realisierungstechniken zuzuordnen sind:

- a) Faltung von Zahlenfolgen mit Zahlenfolgen
- b) Faltung von Zahlenfolgen mit Bit-Folgen
- c) Faltung von Bit-Folgen mit Bit-Folgen.

Die Darstellung a) bedeutet die gewöhnliche Konzentrierte Arithmetik, erfordert also für die Realisierung konventionelle Multiplizierer und Addierer. Das gegenteilige Extrem, Fall c), arbeitet ausschließlich auf der Ebene von Bits: Multiplikationen werden zu logischen UND-Verknüpfungen. Ein Beispiel dafür ist die Zählmethode von Zohar [20,21] zur Realisierung digitaler Filter, eine Abwandlung der Verteilten Arithmetik.

Die Verteilte Arithmetik, wie sie hier verstanden wird, ist durch die Betrachtung b) charakterisiert: Sie vermeidet - ebenso wie c) - die Verwendung von Multiplizierern und ermöglicht statt dessen den günstigen Einsatz von ROMs oder RAMs zur Ab-speicherung von Koeffizientenkombinationen.

Über die drei erwähnten grundsätzlichen Betrachtungs-Ebenen hinaus sind natürlich verschiedene Variationen möglich: z.B. im Fall b) die Zusammenfassung von verschiedenen Blöcken von zu verarbeitenden Wörtern - ein Beispiel hierfür ist die Summen-Zerlegung bei der Realisierung nichtrekursiver Filter - oder auch die Bildung von Bit-Blöcken. Diese Abwandlung enthält die schon zitierte Arbeit [4], in der die Variablen in 2-Bit-Gruppen zerlegt werden, um bei entsprechendem Speicherplatz-Mehraufwand die Geschwindigkeit zu steigern. Auf das Fehlerverhalten hat diese Variante wegen der um 2 Bit versetzten Akkumulation entscheidende Auswirkungen, die im Rahmen dieser Arbeit jedoch nicht betrachtet werden. Es wird lediglich darauf hingewiesen, daß eine Ausweitung der im folgenden entwickelten

Fehleranalyse auf derartige Sonderfälle der Verteilten Arithmetik ohne weiteres möglich ist.

4. FEHLERANALYSE FÜR GLEICHVERTEILTES WEISSES RAUSCHEN

4.1 Prinzipielles zum Fehlerverhalten der Verteilten Arithmetik

4.1.1 Darstellung der verschiedenen Fehlereinflüsse

Es wird zunächst die Berechnung des inneren Produktes eines Koeffizientenvektors $\mathbf{a}$ mit einem Variablenvektor $\mathbf{x}$ in konzentrierter Arithmetik, also unter Verwendung einzelner Multiplizierer und Addierer betrachtet. Die dabei auftretenden Fehler sind Bild 4.1 zu entnehmen.

Die exakt vorliegenden Koeffizienten müssen zunächst auf endliche Wortlängen begrenzt werden, die für die verwendeten Multiplizierer geeignet sind. Das Symbol $[a]_{Q_C}$ beschreibt hier die Rundung der Koeffizienten auf die Quantisierungsstufe Q_C nach der Vorschrift ^{*)}

$$[a]_{Q_C} = \text{sign}(a) \cdot Q_C [Q_C^{-1} |a| + 0.5] .$$

Das gleiche gilt für die Variablen $x_1 \dots x_L$; sie werden auf die Quantisierungsstufe Q_E gerundet, die in der Regel nicht mit Q_C übereinstimmt.

Es wird hier angenommen, daß die Produkte $[a_\rho]_{Q_C} \cdot [x_\rho]_{Q_E}$ zunächst mit unverkürzter Wortlänge addiert werden und erst die Summe auf die Quantisierungsstufe Q_A gerundet wird. Man erhält dann drei verschiedene Fehlereinflüsse am Ausgang, die im folgenden getrennt betrachtet werden:

$\Delta y_C \hat{=}$ Fehler infolge Quantisierung der Koeffizienten

$\Delta y_E \hat{=}$ Fehler infolge Quantisierung der Eingangssignale

$\Delta y_A \hat{=}$ Fehler infolge Quantisierung des Ausgangssignals

Der durch die Begrenzung der Ausgangs-Wortlänge hervorgerufene Fehler Δy_A wird üblicherweise durch eine mit dem Signal y nicht korrelierte Rauschquelle beschrieben und als spektral weiß und

^{*)} $[A] \hat{=}$ größte ganze Zahl, die kleiner oder gleich A ist.

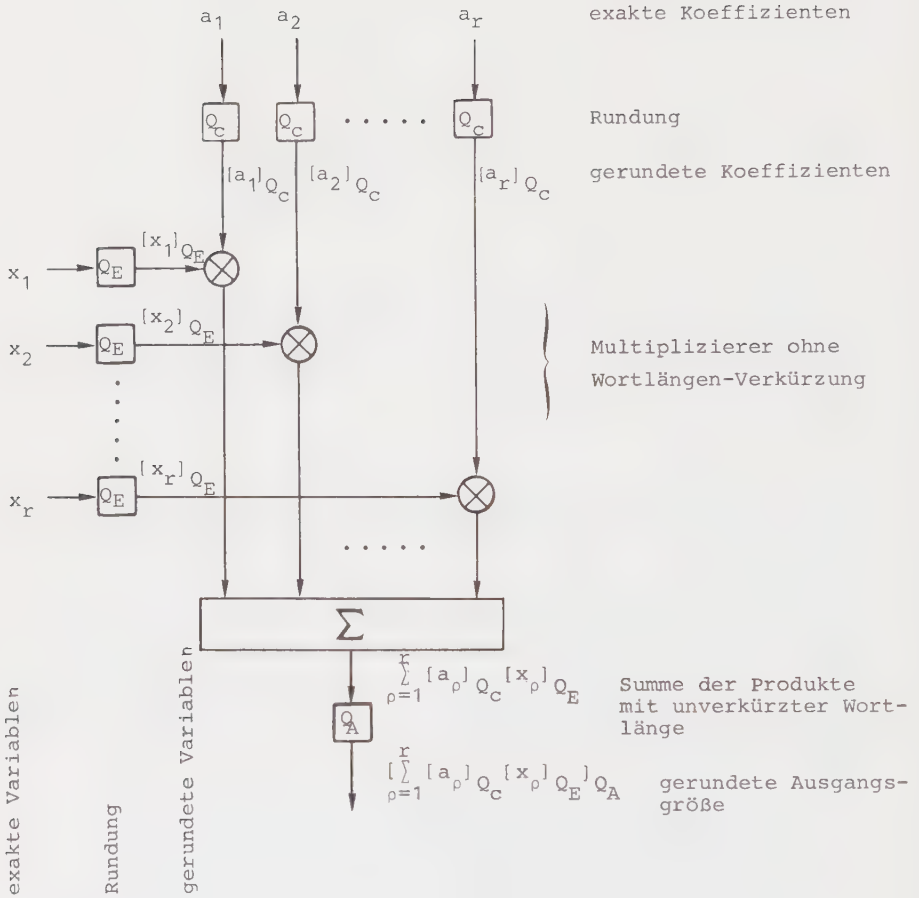


Bild 4.1: Berechnung des inneren Produktes $\mathbf{a}^T \mathbf{x}$ in konzentrierter Arithmetik

$([x]_Q = \text{Größe } x \text{ nach Rundung auf die Quantisierungsstufe } Q)$

im Intervall $[-Q_A/2, +Q_A/2]$ gleichverteilt angenommen (die tatsächliche diskrete Verteilung wird dabei durch eine kontinuierliche angenähert). Seine Leistung beträgt damit:

$$N_{\Delta Y_A} = Q_A^2 / 12 \quad (4.1)$$

Die Quantisierung des Eingangssignals führt am Ausgang zu dem Fehler

$$\Delta y_E = \sum_{\rho=1}^r \Delta x_{\rho} \cdot a_{\rho} \quad (4.2a)$$

mit

$$\Delta x_{\rho} = x_{\rho} - [x_{\rho}]_{Q_E} \quad (4.2b)$$

Unter der Annahme, daß die Fehler Δx_{ρ} jeweils im Intervall $[-\frac{Q_E}{2}, \frac{Q_E}{2}]$ gleichverteilt und untereinander nicht korreliert sind, ergibt sich aus (4.2) die gesamte Fehlerleistung infolge Eingangs-Quantisierung:

$$N_{\Delta Y_E} = \frac{Q_E^2}{12} \sum_{\rho=1}^r a_{\rho}^2 = \frac{Q_E^2}{12} \cdot R_E \quad (4.3)$$

mit der Rauschzahl $R_E = \sum_{\rho=1}^r a_{\rho}^2$.

Die beiden beschriebenen Fehler Δy_A und Δy_E treten ohne Unterschied bei der Verteilten wie bei der Konzentrierten Arithmetik auf. Anders hingegen verhält es sich mit den Auswirkungen der Koeffizientenfehler.

4.1.2 Vergleich der Auswirkung endlicher Koeffizientenspeicherwortlängen bei Konzentrierter und Verteilter Arithmetik

Bei Konzentrierter Arithmetik läßt sich dieser Fehler unmittelbar hinschreiben:

$$\begin{aligned} \Delta y_C &= \mathbf{a}^T \mathbf{x} - [\mathbf{a}^T]_{Q_C} \cdot \mathbf{x} = \\ &= \sum_{\rho=1}^r (a_{\rho} - [a_{\rho}]_{Q_C}) x_{\rho} \end{aligned} \quad (4.4)$$

Mit der Definition der einzelnen Koeffizientenfehler

$$\alpha_{\rho} = a_{\rho} - [a_{\rho}]_{Q_C}, \quad (4.5)$$

sie sollen in dieser Arbeit auch als "a-priori-Koeffizientenfehler" bezeichnet werden, erhält man

$$\Delta y_C = \sum_{\rho=1}^r \alpha_{\rho} x_{\rho}. \quad (4.6)$$

Der Fehler infolge Koeffizientenquantisierung ist also linear von den Eingangssignalen abhängig und führt deshalb ausschließlich zu linearen Verzerrungen des zu realisierenden Systems.

In Kap.3.1 wurde die Schaltung eines flexiblen digitalen Filters in Verteilter Arithmetik angegeben (Bild 3.2), die eine direkte Eingabe der dabei natürlich quantisierten Koeffizienten erlaubt. Für diesen Fall ergeben sich bezüglich der Auswirkung von Koeffizientenfehlern keine Unterschiede gegenüber der Konzentrierten Arithmetik [22].

Im Gegensatz dazu betrachte man nun die Realisierung der Funktion $u(\xi_1, \dots, \xi_r)$ durch ein ROM oder RAM. Hier kann man zur Berechnung des Tafelinhalts zunächst die exakten Koeffizienten verwenden. Erst nach Bildung der verschiedenen Koeffizientenkombinationen werden die Werte auf die Quantisierungsstufe des ROM- bzw. RAM-Ausgangs Q_R gerundet. Dadurch entsteht ein Fehler, der nicht mehr wie bei der Konzentrierten Arithmetik als Folge individueller Koeffizientenfehler interpretiert werden kann. Zur Verdeutlichung sei ein Beispiel angeführt:

Im Akkumulationsschritt i soll der Tafel der Funktionswert $u_i(1,0,1,1) = a_1 + a_3 + a_4$ entnommen werden. Durch endliche Wortlänge des Speichers (Quantisierungsstufe Q_R) enthält der abgespeicherte Wert den Fehler

$$\Delta u_i(1,0,1,1) = (a_1 + a_3 + a_4) - [a_1 + a_3 + a_4]_{Q_R},$$

der nicht mit der Summe der drei einzelnen a-priori Koeffizientenfehler $(\alpha_1 + \alpha_3 + \alpha_4)$ übereinstimmen muß. Das ist schon deshalb unmittelbar einzusehen, weil für den Tafelfehler die Beschrän-

kung $-\frac{Q_R}{2} \leq \Delta u_i \leq \frac{Q_R}{2}$ gilt, was für die Summe der drei Koeffizientenfehler ($-\frac{3}{2} Q_C \leq \alpha_1 + \alpha_3 + \alpha_4 \leq \frac{3}{2} Q_C$) sicher nicht allgemein zutreffen kann.

Den resultierenden Ausgangsfehler infolge Tafelquantisierung erhält man dann durch eine bitweise verschobene Akkumulation von Einzelfehlern Δu_i :

$$\Delta y_O = -\Delta u_O + \sum_{i=1}^{W-1} \Delta u_i 2^{-i} \quad (4.10a)$$

mit

$$\Delta u_i = u(\xi_{1i}, \dots, \xi_{ri}) - [u(\xi_{1i}, \dots, \xi_{ri})]_{Q_R} \quad (4.10b)$$

Der resultierende Fehler in Gl. (4.10a) ist mit einer Null indiziert, was auf den Fall der normalen Zweierkomplement-Codierung der Eingangssignale hinweisen soll. Entsprechend wird der Fehler bei modifizierter Zweierkomplement-Codierung mit Δy_1 bezeichnet. Er ist unter Benutzung von Gl. (2.12)

$$\Delta y_1 = -\frac{1}{2} \Delta v_O + \frac{1}{2} \sum_{i=1}^{W-1} \Delta v_i \cdot 2^{-i} - 2^{-W} \Delta v(1, \dots, 1) \quad (4.11a)$$

mit

$$\Delta v_i = v(\xi_{1i}, \dots, \xi_{ri}) - [v(\xi_{1i}, \dots, \xi_{ri})]_{Q_R} \quad (4.11b)$$

Aus den bis hierher durchgeführten allgemeinen Überlegungen über den Fehler infolge Tafelquantisierung ergeben sich folgende Vermutungen:

Bei hinreichend großer Anzahl von Koeffizienten können die Fehlerwerte Δu_i bzw. Δv_i als statistische Variable angesehen werden, die mit einzelnen Eingangsvariablen nicht mehr korreliert sind. Dadurch entsteht am Ausgang ein Fehler, der durch eine unabhängige Rauschquelle beschrieben werden kann, während einzelne Koeffizientenfehler durch einen "statistischen Mittelungseffekt" verschwinden oder zumindest reduziert werden [23,14].

Für kleine Anzahl von Koeffizienten ist diese statistische Betrachtungsweise sicher nicht zulässig, so daß mit kleiner werdendem r eine sich verstärkende Korrelation des Ausgangsfehlers mit den Eingangssignalen zu erwarten ist. Der Extremfall $r = 1$,

d.h. die Konzentrierte Arithmetik, führt dementsprechend zu ausschließlich linearen Verzerrungen.

Eine genaue Analyse dieser Zusammenhänge, insbesondere auch der Vergleich zwischen normalem und modifiziertem Zweierkomplement, ist das Ziel der folgenden Abschnitte.

4.2 Einfaches Rauschmodell

4.2.1 Fehlerleistung nach Peled und Liu

In [5] wurde zur Berechnung des Fehlers infolge endlicher Tafelwortlänge ein vereinfachtes Modell angegeben. Dabei wurde vorausgesetzt, daß die Eingangsgrößen aus spektral weißen, im Intervall $[-1,1]$ gleichmäßig verteilten Rauschsignalen bestehen, die untereinander nicht korreliert sind. Daraus ergeben sich folgende Bedingungen:

- 1.) Jede der 2^r möglichen Kombinationen der Bits $\xi_{1i}, \dots, \xi_{ri}$ tritt mit der gleichen Wahrscheinlichkeit 2^{-r} auf.
- 2.) Die Bit-Kombination $\xi_{1i}, \dots, \xi_{ri}$ im Akkumulationsschritt i ist statistisch unabhängig von der Kombination $\xi_{1j}, \dots, \xi_{rj}$ eines vorherigen oder nachfolgenden Schrittes $j \neq i$. Es sind also auch die Größen Δu_i und Δu_j für $i \neq j$ statistisch unabhängig.

Zunächst ist unmittelbar zu zeigen, daß der Mittelwert des resultierenden Fehlers $\langle \Delta y_0 \rangle$ *) annähernd verschwindet. Nach Gl. (4.10a) gilt

$$\langle \Delta y_0 \rangle = - \langle \Delta u_0 \rangle + \sum_{i=1}^{w-1} \langle \Delta u_i \rangle 2^{-i} \quad (4.12a)$$

*)

$\langle \Delta y_0 \rangle$ beschreibt den zeitlichen Mittelwert der Folge $\{\Delta y_0(k)\}$:

$$\langle \Delta y_0 \rangle = \lim_{K \rightarrow \infty} \frac{1}{K+1} \sum_{k=0}^K \Delta y_0(k)$$

Die Mittelwerte von Δu_i sind wegen Voraussetzung 1.) für jeden Akkumulationsschritt i gleich, so daß aus (4.12a) folgt:

$$\langle \Delta y_O \rangle = \langle \Delta u \rangle \left(-1 + \sum_{i=1}^{w-1} 2^{-i} \right) \quad (4.12b)$$

Für hinreichend große Wortlängen w wird die Summe $\sum_{i=1}^{w-1} 2^{-i} \approx 1$, woraus mit Gl. (4.12b) die Mittelwertfreiheit des Fehlers Δy_O folgt. Seine Leistung $N_{\Delta y_O}$ ist also mit seiner Varianz $\sigma_{\Delta y_O}^2$ beschrieben:

$$N_{\Delta y_O} = \sigma_{\Delta y_O}^2 + \langle \Delta y_O \rangle^2 \approx \sigma_{\Delta y_O}^2 \quad (4.13)$$

Wegen der Bedingung 2.) kann die Varianz des Gesamtfehlers durch Aufsummation der in jedem Akkumulationsschritt auftretenden Teilvarianzen gewonnen werden. Gl. (4.10a) liefert

$$\sigma_{\Delta y_O}^2 = \sum_{i=0}^{w-1} \sigma_{\Delta u_i}^2 2^{-2i}. \quad (4.14)$$

Mit der Voraussetzung 1.) sind die einzelnen Varianzen $\sigma_{\Delta u_i}^2$ in jedem Akkumulationsschritt i gleich, so daß für die resultierende Ausgangs-Fehlerleistung folgt

$$N_{\Delta y_O} \approx \sigma_{\Delta u_i}^2 \cdot \sum_{i=0}^{w-1} 2^{-2i} \quad (4.14b)$$

Für hinreichend große Eingangswortlängen w ergibt sich schließlich wegen $1 \gg 2^{-2w}$

$$N_{\Delta y_O} \approx \frac{4}{3} \cdot \sigma_{\Delta u}^2. \quad (4.15)$$

In dem Rauschmodell nach Peled und Liu wird nun vereinfachend für Δu eine im Intervall $\left[-\frac{Q_R}{2}, \frac{Q_R}{2} \right]$ gleichmäßig verteilte Zufallsvariable angenommen. Ihre Varianz ist damit

$$\sigma_{\Delta u}^2 = Q_R^2 / 12, \quad (4.16)$$

und man erhält für die resultierende Fehlerleistung infolge Ta-felquantisierung den Wert

$$N_{\Delta y_0} \approx \frac{Q_R^2}{9} = 0.1111 Q_R^2 \quad (4.17)$$

Eine Erweiterung dieses Verfahrens auf den Fall des modifizierten Zweierkomplements ist leicht durchzuführen:

Sieht man von der vernachlässigbaren Gleichgröße $2^{-w} \Delta v(1, \dots, 1)$ des Ausgangsfehlers Δy_1 ab, so ist die resultierende Fehlerleistung wieder durch Aufsummation der Einzelvarianzen in den verschiedenen Akkumulationsschritten zu berechnen (vgl. Gl. 4.11a):

$$N_{\Delta y_1} = \sum_{i=0}^{w-1} \sigma_{\Delta v_i}^2 2^{-2i-2} \quad (4.18)$$

Die Werte $\sigma_{\Delta v_i}^2$ sind auch hier wegen der Voraussetzung 1.) alle gleich, so daß sich ergibt

$$N_{\Delta y_1} = \sigma_{\Delta v}^2 \cdot \frac{1}{4} \sum_{i=0}^{w-1} 2^{-2i} = \frac{1}{3} \cdot \sigma_{\Delta v}^2 \quad (4.19)$$

Unter der Annahme einer Gleichverteilung für Δv im Intervall $\left[-\frac{Q_R}{2}, \frac{Q_R}{2}\right]$ erhält man für die resultierende Fehlerleistung bei der modifizierten Zweierkomplement-Codierung

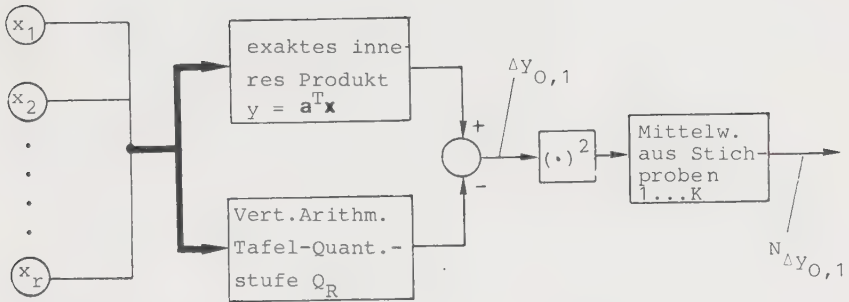
$$N_{\Delta y_1} = \frac{1}{36} Q_R^2 = 0.02777 \cdot Q_R^2 \quad (4.20)$$

4.2.2 Simulation einiger Beispiele *)

Die Ergebnisse (4.17) und (4.20) sollen durch Simulation entsprechend dem Blockdiagramm in Bild 4.2 überprüft werden.

Zur Simulation werden drei Beispiele mit zufällig gewählten Koeffizientensätzen benutzt. Die Eingangswortlänge ist dabei jeweils $w = 8$, die Tafel-Quantisierungsstufe $Q_R = 2^{-7}$. In allen Fällen wird eine Stichprobenmenge von $K = 50\,000$ für die Leistungsberechnung angesetzt.

*) Sämtliche in der Arbeit behandelten Beispiele sind in Anhang 10 tabellarisch zusammengestellt.



r unabhängige, weiße
Rauschquellen, gleich-
verteilt in $[-1, 1]$
(Stichproben 1...K)

Bild 4.2: Blockdiagramm zur Berechnung der Fehlerleistung
durch Simulation

- Beispiel 1: $r = 2$

$$a_1 = 0.2178328 \rightarrow \alpha_1 = -0.1174 Q_R$$

$$a_2 = -0.3159031 \rightarrow \alpha_2 = -0.4356 Q_R$$

Simulationsergebnis

$$N_{\Delta y_O} = 0.1331 Q_R^2$$

$$N_{\Delta y_1} = 0.0504 Q_R^2$$

- Beispiel 2: $r = 2$

$$a_1 = 0.4941047 \rightarrow \alpha_1 = 0.2454 Q_R$$

$$a_2 = 0.1852891 \rightarrow \alpha_2 = -0.2830 Q_R$$

Simulationsergebnis

$$N_{\Delta y_O} = 0.04669 Q_R^2$$

$$N_{\Delta y_1} = 0.03735 Q_R^2$$

- Beispiel 3: $r = 8$

$$a_1 = 0.251531 \quad \rightarrow \quad \alpha_1 = 0.196 \quad Q_R$$

$$a_2 = 0.218546 \quad \rightarrow \quad \alpha_2 = -0.026 \quad Q_R$$

$$a_3 = -0.405625 \quad \rightarrow \quad \alpha_3 = 0.080 \quad Q_R$$

$$a_4 = 0.849391 \quad \rightarrow \quad \alpha_4 = -0.278 \quad Q_R$$

$$a_5 = -0.260927 \quad \rightarrow \quad \alpha_5 = -0.399 \quad Q_R$$

$$a_6 = 0.282398 \quad \rightarrow \quad \alpha_6 = 0.147 \quad Q_R$$

$$a_7 = 0.081344 \quad \rightarrow \quad \alpha_7 = 0.412 \quad Q_R$$

$$a_8 = 0.155047 \quad \rightarrow \quad \alpha_8 = 0.154 \quad Q_R$$

$$\begin{aligned} \text{Simulationsergebnis} \quad N_{\Delta Y_0} &= 0.1079 \quad Q_R^2 \\ N_{\Delta Y_1} &= 0.0282 \quad Q_R^2 \end{aligned}$$

Man erkennt, daß die Fehlerleistungen der beiden Beispiele für 2 Variable stark voneinander abweichen und zwar sowohl bei der Zweierkomplement-Codierung als auch beim modifizierten Zweierkomplement. Mit dem Wert $Q_R^2/9$ bzw. $Q_R^2/36$ stimmt für $r = 2$ keines der Simulationsergebnisse überein. Es ist also offenbar bei kleiner Koeffizientenzahl nicht möglich, den Fehler zu formulieren, ohne die individuellen Systemparameter zu berücksichtigen. Das heißt die Annahme der Gleichverteilung für Δu bzw. Δv unabhängig von der Anzahl der Koeffizienten und den Koeffizientenwerten selbst ist eine zu grobe Vereinfachung, so daß dieses Modell im Falle $r = 2$ völlig unbrauchbare Ergebnisse liefert.

Die Bilder 4.3a,b und 4.4a,b zeigen die wirklichen Verteilungsdichte-Funktionen für die Beispiele 1 und 2. *) Naturgemäß sind sie diskret wegen der endlichen Anzahl von $2^r = 4$ möglichen Werten für $u(\xi_1, \xi_2)$ bzw. $v(\xi_1, \xi_2)$.

*)

$$p_{\Delta u}(\beta) = \frac{P(\beta \leq \Delta u < \beta + \Delta\beta)}{\Delta\beta}$$

$$p_{\Delta v}(\beta) = \frac{P(\beta \leq \Delta v < \beta + \Delta\beta)}{\Delta\beta}$$

$$\Delta\beta = \frac{Q}{R}$$

Die Verteilungen von Δu und Δv für das Beispiel von 8 Variablen (Bild 4.5a, 4.5b) hingegen kommen der angenommenen Gleichverteilung im Intervall $\left[-\frac{Q_R}{2}, \frac{Q_R}{2}\right]$ näher. Das Simulationsergebnis stimmt daher mit den Werten nach (4.17) bzw. (4.20) annähernd überein.

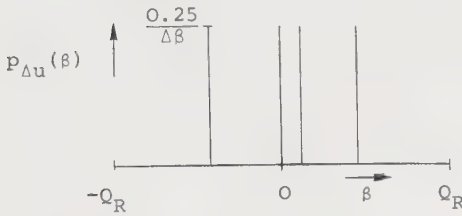


Bild 4.3a: Zweierkomplement
r = 2
Beispiel 1

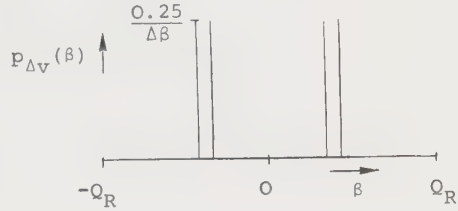


Bild 4.3b: Modifizierte Form
r = 2
Beispiel 1

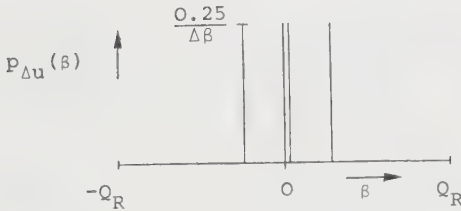


Bild 4.4a: Zweierkomplement
r = 2
Beispiel 2

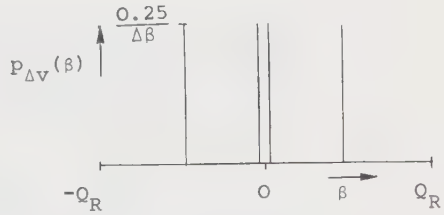


Bild 4.4b: Modifizierte Form
r = 2
Beispiel 2

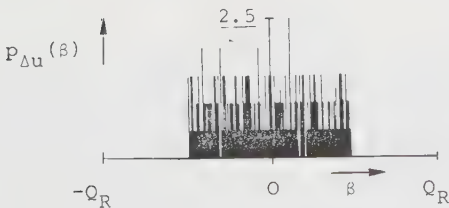


Bild 4.5a: Zweierkomplement
r = 8
Beispiel 3

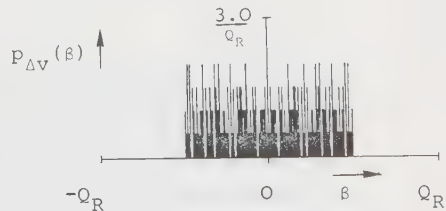


Bild 4.5b: Modifizierte Form
r = 8
Beispiel 3

4.2.3 Berechnung der exakten Fehlervarianzen für individuelle Systeme

Zu einer exakten Formulierung des Ausgangsfehlers bei weißen, gleichverteilten und unabhängigen Rauschsignalen am Eingang kommt man, wenn für $\sigma_{\Delta u}^2$ bzw. $\sigma_{\Delta v}^2$ die für den speziellen Koeffizientensatz korrekten Werte gesetzt werden [24].

Sie sind jeweils aus den Verteilungen von Δu bzw. Δv zu ermitteln:

$$\sigma_{\Delta u}^2 = \sum_{\lambda=1}^{2^r} \left\{ \Delta u(\lambda) - \sum_{\kappa=1}^{2^r} (\Delta u(\kappa) p_{\kappa}) \right\}^2 \cdot p_{\lambda} \quad (4.21a)$$

$$\sigma_{\Delta v}^2 = \sum_{\lambda=1}^{2^r} \left\{ \Delta v(\lambda) - \sum_{\kappa=1}^{2^r} (\Delta v(\kappa) p_{\kappa}) \right\}^2 \cdot p_{\lambda} \quad (4.21b)$$

Hierbei bedeutet $\Delta u(\lambda)$ bzw. $\Delta v(\lambda)$ je einen der 2^r möglichen Werte. Die zugehörige Wahrscheinlichkeit ist p_{λ} , die wegen Voraussetzung 1.) für jedes λ den Wert 2^{-r} hat. Bei der modifizierten Form ist darüber hinaus zu berücksichtigen, daß der Mittelwert $\langle \Delta v_1 \rangle = \sum_{\lambda=1}^{2^r} \Delta v(\lambda) p_{\lambda}$ verschwindet, weil zu jedem Wert $v(\xi_{1,i}, \dots, \xi_{r,i})$ ein gleicher Wert mit umgekehrtem Vorzeichen gehört (vgl. Abschnitt 2.2). Aus den Gleichungen (4.21) ergibt sich damit

$$\sigma_{\Delta u}^2 = 2^{-r} \sum_{\lambda=1}^{2^r} \left\{ \Delta u(\lambda) - 2^{-r} \sum_{\kappa=1}^{2^r} \Delta u(\kappa) \right\}^2 \quad (4.22a)$$

$$\sigma_{\Delta v}^2 = 2^{-r} \sum_{\lambda=1}^{2^r} (\Delta v(\lambda))^2 \quad (4.22b)$$

Die Gleichungen (4.22a,b) sind für ein gegebenes Beispiel leicht zu programmieren: Es müssen alle möglichen Koeffizientenkombinationen gebildet und die jeweils zugehörigen Quantisierungsfehler berechnet werden, also

$$\begin{aligned} \Delta u(\lambda=1) &= 0 \\ \Delta u(\lambda=2) &= a_1 - [a_1]_{Q_R} \\ \Delta u(\lambda=3) &= a_2 - [a_2]_{Q_R} \\ \Delta u(\lambda=4) &= (a_1 + a_2) - [a_1 + a_2]_{Q_R} \\ &\vdots \end{aligned}$$

$$\begin{aligned}
 \text{bzw.} \quad \Delta v(\lambda=1) &= (a_1 + a_2 + \dots + a_r) - [a_1 + a_2 + \dots + a_r]_{Q_R} \\
 \Delta v(\lambda=2) &= (-a_1 + a_2 + \dots + a_r) - [-a_1 + a_2 + \dots + a_r]_{Q_R} \\
 \Delta v(\lambda=3) &= (+a_1 - a_2 + \dots + a_r) - [+a_1 - a_2 + \dots + a_r]_{Q_R} \\
 &\vdots
 \end{aligned}$$

Die anschließende Aufsummation gemäß Gl. (4.22a,b) liefert die Varianz.

Für die oben betrachteten Beispiele erhält man nach dieser Methode:

$$\text{Beispiel 1:} \quad \sigma_{\Delta u}^2 = 0.10013 Q_R^2 ; \quad \sigma_{\Delta v}^2 = 0.15054 Q_R^2$$

$$\text{Beispiel 2:} \quad \sigma_{\Delta u}^2 = 0.03508 Q_R^2 ; \quad \sigma_v^2 = 0.11190 Q_R^2$$

$$\text{Beispiel 3:} \quad \sigma_{\Delta u}^2 = 0.08092 Q_R^2 ; \quad \sigma_{\Delta v}^2 = 0.08458 Q_R^2$$

Nach Einsetzen in Gl. (4.15) und (4.19) ergeben sich die resultierenden Fehlerleistungen

$$\text{Beispiel 1:} \quad N_{\Delta y_O} = 0.13351 Q_R^2 ; \quad N_{\Delta y_1} = 0.05018 Q_R^2$$

$$\text{Beispiel 2:} \quad N_{\Delta y_O} = 0.04677 Q_R^2 ; \quad N_{\Delta y_1} = 0.03723 Q_R^2$$

$$\text{Beispiel 3:} \quad N_{\Delta y_O} = 0.10789 Q_R^2 ; \quad N_{\Delta y_1} = 0.02819 Q_R^2$$

die mit den oben angegebenen Simulationsergebnissen sehr genau übereinstimmen.

Mit dieser Methode ist somit die exakte Berechnung der Ausgangs-Fehlerleistung möglich für den Fall, daß am Eingang r unabhängige, spektral weiße und gleichverteilte Rauschsignale liegen. Damit ist allerdings keine Aussage darüber zu machen, inwieweit der vermutete "statistische Mittelungseffekt" über einzelne Koeffizientenfehler stattgefunden hat, denn etwaige verbleibende restliche Koeffizientenfehler führen am Ausgang wegen der rauschartigen Erregung ebenfalls zu einem weißen, gleichverteilten Fehler, der von dem unkorrelierten Anteil nach dieser Methode nicht zu trennen ist.

Ferner ist die Ausweitung dieses Modells auf farbige bzw. nicht gleichverteilte Eingangssignale nicht ohne weiteres möglich. Es wird daher im nächsten Abschnitt ein allgemeiner Ansatz durchgeführt, der neben einer geschlossenen analytischen Formulierung der Gesamtfehlerleistung die Trennung nach linear mit den Eingangssignalen korrelierten und unkorrelierten Termen erlaubt. Die Erweiterung auf allgemeinere Eingangssignaleigenschaften ist dabei möglich und wird in späteren Kapiteln durchgeführt.

4.3 Allgemeine Formulierung des Ausgangsfehlers infolge endlicher Tafelwortlänge *)

4.3.1 Zweierkomplement-Codierung der Eingangssignale

Die Quantisierungsfehler der einzelnen abgespeicherten Tafelwerte lassen sich mit Hilfe der in Gl. (4.5) definierten individuellen a-priori-Koeffizientenfehler formulieren, wenn $Q_R = Q_C$ gesetzt wird.

$$a_\rho = [a_\rho]_{Q_R} + \alpha_\rho \quad (4.23)$$

Für Δu_i erhält man nach Einsetzen von (4.23) in (4.10b):

$$\Delta u_i = \sum_{\rho=1}^r \alpha_\rho \xi_{\rho i} + \sum_{\rho=1}^r [a_\rho]_{Q_R} \xi_{\rho i} - \left[\sum_{\rho=1}^r a_\rho \xi_{\rho i} \right]_{Q_R} \quad (4.24)$$

Der zweite Term in Gl. (4.24) besteht aus einer Summe von ausschließlich gerundeten Werten. Er kann deshalb mit dem dritten Summenausdruck zusammengefaßt werden, weil dessen Rundungsoperation dadurch nicht beeinflußt wird.

$$\begin{aligned} \sum_{\rho=1}^r [a_\rho]_{Q_R} \xi_{\rho i} - \left[\sum_{\rho=1}^r a_\rho \xi_{\rho i} \right]_{Q_R} &= - \left[\sum_{\rho=1}^r (a_\rho - [a_\rho]_{Q_R}) \xi_{\rho i} \right]_{Q_R} = \\ &= - \left[\sum_{\rho=1}^r \alpha_\rho \xi_{\rho i} \right]_{Q_R} \end{aligned} \quad (4.25)$$

*) Die folgende Fehleranalyse für weißes, gleichverteiltes Rauschen wurde für den Fall der Zweierkomplement-Codierung zur Veröffentlichung eingereicht [25].

Aus den Gleichungen (4.24) und (4.25) erhält man einen Ausdruck für Δu_i , der nicht mehr von den Koeffizienten selbst, sondern nur noch von den a-priori-Koeffizientenfehlern α_ρ abhängt.

$$\Delta u_i = \sum_{\rho=1}^r \alpha_\rho \xi_{\rho i} - Q_R \left(\frac{1}{Q_R} \left[\sum_{\rho=1}^r \alpha_\rho \xi_{\rho i} \right]_{Q_R} \right) \quad (4.26)$$

Der Term $\frac{1}{Q_R} \left[\sum_{\rho=1}^r \alpha_\rho \xi_{\rho i} \right]_{Q_R}$ in Gl. (4.26) kann für alle 2^r Kombinationen der Bits $\xi_{1i}, \dots, \xi_{ri}$ nur ganzzahlige Werte annehmen. Um zu einer geschlossenen Formulierung zu kommen, wird jetzt ein Fehlervektor $\mathbf{e}^{(0)}$ eingeführt, der aus 2^r Komponenten besteht.

$$\mathbf{e}^{(0)} = (e_0^{(0)}, e_1^{(0)}, \dots, e_n^{(0)}, \dots, e_{2^r-1}^{(0)})^T \quad (4.27)$$

Der hochgestellte Index (0) kennzeichnet den Fall der einfachen Zweierkomplement-Codierung der Eingangssignale.

Die einzelnen Komponenten sind nach folgender Vorschrift zu bilden:

$$e_n^{(0)} = \left(- \frac{1}{Q_R} \left[\sum_{\rho=1}^r \alpha_\rho v_\rho \right]_{Q_R} \right); e_n \in \mathbb{Z}, \quad (4.28a)$$

wobei die Größen v_ρ die Bits des gemäß Gl. (4.28b) binär dargestellten Komponenten-Index' n bedeuten.

$$n = \sum_{\rho=1}^r v_\rho 2^{\rho-1} \quad 0 \leq n \leq 2^r-1, v_\rho \in \{0,1\} \quad (4.28b)$$

Die im Fehlervektor enthaltene Information kann anschaulich gedeutet werden: Ist $e_n^{(0)} = 0$, so bedeutet das nach Gl. (4.28a) offenbar, daß die Summe der zugehörigen a-priori-Koeffizientenfehler kleiner als $Q_R/2$ ist. Für $e_n^{(0)} = \pm 1$ hingegen wird die Quantisierungsstufe von der Koeffizientenkombination einmal übersprungen. Der entsprechende Fehler Δu_i ist dann nicht mehr gleich der Summe der a-priori-Koeffizientenfehler wie bei $e_n^{(0)} = 0$, sondern nach Gl. (4.26) ist der Wert Q_R zu subtrahieren ($e_n^{(0)} = -1$) bzw. zu addieren ($e_n^{(0)} = +1$). Für mehr als drei Variable kann $|e_n^{(0)}|$ auch größere Werte als 1 annehmen, d.h. in dem Fall wird die Quantisierungsstufe mehrfach übersprungen:

Der Wert der Fehlervektor-Komponente $e_n^{(0)}$ gibt also an, wie oft in der zugehörigen Kombination der a-priori-Koeffizientenfehler eine Quantisierungsstufe übersprungen wird. Das kann bei r Koeffizienten $\left[\frac{r}{2}\right]$ mal passieren, wobei $\left[\frac{r}{2}\right]$ die größte ganze Zahl bedeutet, die kleiner oder gleich $\frac{r}{2}$ ist. Es gilt damit

$$- \left[\frac{r}{2}\right] \leq e_n^{(0)} \leq \left[\frac{r}{2}\right], \quad e_n \in \mathbb{Z} \quad (4.29)$$

Mit Hilfe des Fehlervektors $e^{(0)}$ kann nun das Fehlerverhalten von individuellen Systemen wie folgt beschrieben werden. Zunächst wird derjenige Index n ausgewählt, dessen r Bits v_ρ der Binärdarstellung nach Gl. (4.28b) mit den aktuellen Eingangsbits $\xi_{1i}, \dots, \xi_{ri}$ übereinstimmen. Das ist formal hinzuschreiben, indem man das Produkt aller r Valenzoperationen *) von $\xi_{\rho i}$ und v_ρ bildet, das nur dann verschieden von Null ist, wenn alle Bit-Paare gleich sind.

$$\prod_{\rho=1}^r (\xi_{\rho i} \ominus v_\rho) = \begin{cases} 1 & \text{für } \xi_{\rho i} = v_\rho \quad \forall \rho=1, \dots, r \\ 0 & \text{sonst.} \end{cases} \quad (4.30)$$

Ausgehend von Gleichung (4.26) lautet die Formulierung des Einzelfehlers Δu_i dann mit Hilfe des Fehlervektors:

$$\Delta u_i = \sum_{\rho=1}^r \alpha_\rho \xi_{\rho i} + Q_R \sum_{n=0}^{2^r-1} (e_n^{(0)} \cdot \prod_{\rho=1}^r (\xi_{\rho i} \ominus v_\rho)) \quad (4.31)$$

Diese Gleichung wird in (4.10a) eingesetzt und liefert schließlich einen Ausdruck für den resultierenden Ausgangsfehler (vgl. Anhang 1):

$$\Delta y_0 = \sum_{\rho=1}^r \alpha_\rho x_\rho + Q_R \sum_{n=0}^{2^r-1} e_n^{(0)} \cdot B(n, \mathbf{x}) \quad (4.32)$$

Die vom Index n und dem Eingangsvektor $\mathbf{x}$ abhängige Funktion

$$B(n, \mathbf{x}) = - \prod_{\rho=1}^r (\xi_{\rho 0} \ominus v_\rho) + \sum_{i=1}^{w-1} (2^{-i} \prod_{\rho=1}^r (\xi_{\rho i} \ominus v_\rho)) \quad (4.33)$$

bildet dabei jeweils das Gewicht der einzelnen Fehlervektorkomponenten $e_n^{(0)}$.

*)

$$\xi_{\rho i} \ominus v_\rho = \begin{cases} 1 & \text{für } \xi_{\rho i} = v_\rho \\ 0 & \text{für } \xi_{\rho i} = \bar{v}_\rho \end{cases} \quad \hat{=} \text{Valenzoperation}$$

An dieser Stelle ist die Notwendigkeit einer derartig umständlich anmutenden Formulierung des Fehlers nicht recht einzusehen, zumal in Gl. (4.10a) schon ein Ausdruck zur korrekten Berechnung von Δy_0 vorliegt. Um den Sinn der Darstellung nach Gl. (4.32) zu verdeutlichen, soll im folgenden ein Beispiel für den Fall $r = 3$ durchgeführt werden. Die zufällig gewählten Koeffizienten und die zugehörigen a-priori-Koeffizientenfehler (für $Q_R = 2^{-7}$) sind:

$$\begin{array}{ll} a_1 = 0.31586 & \alpha_1 = 0.43008 Q_R \\ a_2 = -0.18507 & \alpha_2 = 0.31104 Q_R \\ a_3 = 0.37320 & \alpha_3 = -0.23040 Q_R \end{array}$$

Zunächst wird der Ausgangsfehler nach der direkten Berechnungsmethode (4.10a) bestimmt. Dabei sind die 8 möglichen Quantisierungsfehler der in der Tafel abgespeicherten Werte $u(\xi_{1i}, \xi_{2i}, \xi_{3i})$ nach Gl. (4.10b) zu berechnen.

ξ_{1i}	ξ_{2i}	ξ_{3i}	$\Delta u(\xi_{1i}, \xi_{2i}, \xi_{3i})$
0	0	0	0
0	0	1	-0.23040 Q_R
0	1	0	0.31104 Q_R
0	1	1	0.08064 Q_R
1	0	0	0.43008 Q_R
1	0	1	0.19968 Q_R
1	1	0	-0.25888 Q_R
1	1	1	-0.48928 Q_R

Der aktuelle Zustand des Eingangsvektors $\mathbf{x}$ möge durch die Komponenten

$$\left. \begin{array}{lll} x_1 = 0.7265625 & \hat{=} & 0.1011101 \\ x_2 = -0.8203125 & \hat{=} & 1.0010111 \\ x_3 = 0.2031250 & \hat{=} & 0.0011010 \end{array} \right\} \begin{array}{l} \text{Zweierkomplement-} \\ \text{Darstellung} \end{array}$$

festgelegt sein. Dann erhält man mit Gl. (4.10a) den resultierenden Ausgangsfehler

$$\begin{aligned}\Delta y_0 &= -\Delta u(0,1,0) + 2^{-1} \Delta u(1,00) + 2^{-2} \Delta u(0,0,0) + \dots \\ &\dots + 2^{-7} \Delta u(1,1,0) \\ &= -0.1535325 Q_R\end{aligned}$$

Bei der Berechnung von Δy_0 nach den in diesem Abschnitt abgeleiteten Gleichungen wird zuerst der Fehlervektor bestimmt. Für seine Komponenten ergibt sich nach Gl. (4.28a):

$$\begin{aligned}n = 0 & : e_0^{(0)} = 0 \\ n = 1 = (2^0) & : e_1^{(0)} = -\frac{1}{Q_R} [\alpha_1]_{Q_R} = -\frac{1}{Q_R} [0.43008 Q_R]_{Q_R} = 0 \\ n = 2 = (2^1) & : e_2^{(0)} = -\frac{1}{Q_R} [\alpha_2]_{Q_R} = -\frac{1}{Q_R} [0.31104 Q_R]_{Q_R} = 0 \\ n = 3 = (2^0 + 2^1) & : e_3^{(0)} = -\frac{1}{Q_R} [\alpha_1 + \alpha_2]_{Q_R} = -\frac{1}{Q_R} [0.74112 Q_R]_{Q_R} = -1 \\ n = 4 = (2^2) & : e_4^{(0)} = -\frac{1}{Q_R} [\alpha_3]_{Q_R} = -\frac{1}{Q_R} [-0.2304 Q_R]_{Q_R} = 0 \\ n = 5 = (2^0 + 2^2) & : e_5^{(0)} = -\frac{1}{Q_R} [\alpha_1 + \alpha_3]_{Q_R} = -\frac{1}{Q_R} [0.19968 Q_R]_{Q_R} = 0 \\ n = 6 = (2^1 + 2^2) & : e_6^{(0)} = -\frac{1}{Q_R} [\alpha_2 + \alpha_3]_{Q_R} = -\frac{1}{Q_R} [0.08064 Q_R]_{Q_R} = 0 \\ n = 7 = (2^0 + 2^1 + 2^2) & : e_7^{(0)} = -\frac{1}{Q_R} [\alpha_1 + \alpha_2 + \alpha_3]_{Q_R} = -\frac{1}{Q_R} [0.51072 Q_R]_{Q_R} = -1\end{aligned}$$

Für den hier verwendeten Koeffizientensatz erhält man also den Fehlervektor

$$\mathbf{e}^{(0)} = (0, 0, 0, -1, 0, 0, 0, -1)^T.$$

Es gilt damit für jeden beliebigen Eingangsvektor $\mathbf{x}$

$$Q_R \sum_{n=0}^7 e_n^{(0)} B(n, \mathbf{x}) = Q_R (-B(3, \mathbf{x}) - B(7, \mathbf{x})).$$

Andererseits lassen sich unabhängig vom Koeffizientenvektor $\mathbf{a}$ nach Gl. (4.33) die Werte $B(n, \mathbf{x})$ berechnen. Für das oben eingeführte Beispiel des Variablenvektors $\mathbf{x}$ ergibt sich:

n	v ₁	v ₂	v ₃	i	B(n, x)
0	0	0	0	2	2 ⁻²
1	1	0	0	1	2 ⁻¹
2	0	1	0	0	-2 ⁰
3	1	1	0	5, 7	2 ⁻⁵ +2 ⁻⁷
4	0	0	1	-	0
5	1	0	1	4	2 ⁻⁴
6	0	1	1	6	2 ⁻⁶
7	1	1	1	3	2 ⁻³

Für den vorliegenden Koeffizientensatz ist dann

$$Q_R \sum_{n=0}^7 e_n^{(0)} B(n, \mathbf{x}) = -Q_R (2^{-5} + 2^{-7} + 2^{-3}) = -0.1640625 Q_R .$$

Schließlich ergibt sich für den resultierenden Fehler mit Gl. (4.32)

$$\Delta y_O = \sum_{\rho=1}^3 \alpha_{\rho} x_{\rho} - 0.1640625 Q_R = -0.1535325 Q_R ,$$

also der gleiche Wert wie bei der direkten Anwendung der Gleichungen (4.10a,b).

Anhand dieses Beispielles wird die Besonderheit der Fehlerformulierung nach (4.32, 4.33) deutlich: Es wird eine Darstellung des Fehlers auf Bit-Ebene durchgeführt, bei der die Einflüsse des Koeffizientenvektors (→ Fehlervektor) von denen des Eingangsvektors (→ B(n, **x**)) getrennt sind. Es wurde bereits darauf hingewiesen, daß diese Eigenschaft in zweierlei Hinsicht von entscheidender Bedeutung ist: Einmal ermöglicht sie die Berechnung der Korrelation zwischen dem Fehler Δy_O und den Eingangssignalen x_{ρ} , zum anderen erlaubt sie die Fehler-Untersuchung auch bei nicht gleichverteilten Eingangssignalen (Kap. 6).

4.3.2 Modifizierte Zweierkomplement-Codierung der Eingangssignale

Bei Verwendung der modifizierten Zweierkomplement-Codierung ergibt sich aus Gl. (2.11) der Einzelfehler im Akkumulations-schritt i.

$$\begin{aligned}\Delta v_i &= v(\xi_{1i}, \dots, \xi_{ri}) - [v(\xi_{1i}, \dots, \xi_{ri})]_{Q_R} = \\ &= \sum_{\rho=1}^r (\xi_{\rho i} - \bar{\xi}_{\rho i}) a_{\rho} - \left[\sum_{\rho=1}^r (\xi_{\rho i} - \bar{\xi}_{\rho i}) a_{\rho} \right]_{Q_R} \quad (4.34)\end{aligned}$$

Entsprechend der Ableitung in Abschnitt 4.3.1 werden die a-priori-Koeffizientenfehler eingeführt:

$$\begin{aligned}\Delta v_i &= \sum_{\rho=1}^r (\xi_{\rho i} - \bar{\xi}_{\rho i}) \alpha_{\rho} + \sum_{\rho=1}^r (\xi_{\rho i} - \bar{\xi}_{\rho i}) \cdot [a_{\rho}]_{Q_R} - \\ &- \left[\sum_{\rho=1}^r (\xi_{\rho i} - \bar{\xi}_{\rho i}) a_{\rho} \right]_{Q_R} \quad (4.35)\end{aligned}$$

Wiederum können diejenigen Summen zusammengefaßt werden, die Rundungsoperationen enthalten:

$$\Delta v_i = \sum_{\rho=1}^r (\xi_{\rho i} - \bar{\xi}_{\rho i}) \alpha_{\rho} - \left[\sum_{\rho=1}^r (\xi_{\rho i} - \bar{\xi}_{\rho i}) \alpha_{\rho} \right]_{Q_R}, \quad (4.36)$$

und man erhält auch hier einen Ausdruck, der nicht mehr von den Koeffizienten selbst, sondern nur noch von deren Quantisierungsfehlern abhängt.

Gleichung (4.11a) liefert schließlich den resultierenden Fehler infolge endlicher Tafel-Wortlänge für die modifizierte Form. (siehe Anhang 2).

$$\begin{aligned}\Delta y_1 &= \sum_{\rho=1}^r \alpha_{\rho} x_{\rho} + \frac{1}{2} \left[\sum_{\rho=1}^r \alpha_{\rho} (\xi_{\rho 0} - \bar{\xi}_{\rho 0}) \right]_{Q_R} - \frac{1}{2} \sum_{i=1}^{w-1} 2^{-i} \left[\sum_{\rho=1}^r \alpha_{\rho} (\xi_{\rho i} - \bar{\xi}_{\rho i}) \right]_{Q_R} \\ &+ 2^{-w} \left[\sum_{\rho=1}^r \alpha_{\rho} \right]_{Q_R} \quad (4.37)\end{aligned}$$

Man führt nun auch hier einen geeignet gewählten Fehlervektor ein:

$$\mathbf{e}^{(1)} = (e_0^{(1)}, e_1^{(1)}, \dots, e_n^{(1)}, \dots, e_{2^r-1}^{(1)})^T, \quad (4.38a)$$

dessen Komponenten nach (4.38b) definiert sind.

$$e_n^{(1)} = -\frac{1}{2} \cdot \frac{1}{Q_R} \left[\sum_{\rho=1}^r \alpha_\rho (v_\rho - \bar{v}_\rho) \right]_{Q_R}^* . \quad (4.38b)$$

Dann erhält man analog zu Abschnitt 4.3.1 eine Fehlerformulierung auf der Basis dieses Vektors $\mathbf{e}^{(1)}$

$$\Delta y_1 = \sum_{\rho=1}^r \alpha_\rho \mathbf{x}_\rho + Q_R \cdot \sum_{n=0}^{2^r-1} e_n^{(1)} \cdot B(n, \mathbf{x}) - 2^{-(w-1)} Q_R \cdot e_{2^r-1}^{(1)} \quad (4.39a)$$

mit den gleichen Ausdrücken für die Gewichte der Fehlervektor-Komponenten wie bei einfacher Zweierkomplement-Codierung:

$$B(n, \mathbf{x}) = - \prod_{\rho=1}^r (\xi_{\rho 0} \oplus v_\rho) + \sum_{i=1}^{w-1} 2^{-i} \prod_{\rho=1}^r (\xi_{\rho i} \oplus v_\rho) \quad (4.39b)$$

Sieht man von dem Gleichanteil $2^{-(w-1)} Q_R \cdot e_{2^r-1}^{(1)}$ ab, so hat man mit Gl. (4.39) die gleiche Formulierung des resultierenden Fehlers gefunden wie bei der Zweierkomplement-Codierung (vgl. Gl. 4.32). Das unterschiedliche Fehlerverhalten der beiden Realisierungsformen wird also ausschließlich durch die verschieden definierten Fehlervektoren erfaßt. So ist nach der Gleichung (4.38b) bei jeder Komponente $e_n^{(1)}$ der gesamte Koeffizientenfehler-Satz beteiligt - im Gegensatz zum Fehlervektor $\mathbf{e}^{(0)}$. Wegen der Vorzeichensymmetrie gilt für den Vektor $\mathbf{e}^{(1)}$

$$e_{2^r-1-m}^{(1)} = - e_{2^r-1+m-1}^{(1)}, \quad m \in \{1, \dots, 2^r-1\}, \quad (4.40a)$$

so daß bei der modifizierten Form die Summe über alle Fehlervektor-Komponenten immer verschwindet

$$\sum_{n=0}^{2^r-1} e_n^{(1)} = 0, \quad (4.40b)$$

*) Zur Definition von v_ρ vgl. Gl. (4.28b).

was für $e^{(0)}$ nicht zutrifft.

Schließlich besteht noch ein Unterschied im Wertevorrat der beiden Fehlervektoren. Die Komponenten $e_n^{(0)}$ sind grundsätzlich ganze Zahlen, während man gemäß Gl. (4.38b) für $e_n^{(1)}$ Vielfache von $\frac{1}{2}$ erhält:

$$2 \cdot e_n^{(1)} \in \mathbb{Z} \quad (4.40c)$$

Der Vorteil der formal gleichen Darstellung von Δy_0 und Δy_1 durch entsprechende Definition der beiden Fehlervektoren liegt auf der Hand: Alle Ableitungen der folgenden Abschnitte können für Zweierkomplement und die modifizierte Form immer gemeinsam durchgeführt werden. Allerdings muß dabei jeweils der in (4.39a) enthaltene konstante Anteil gesondert betrachtet werden.

4.4 Gesamtleistung des Fehlers

4.4.1 Statistische Eigenschaften der Eingangssignale

Die in den letzten beiden Abschnitten hergeleiteten Beziehungen für die Fehlergrößen gelten allgemein für beliebige Signale x_p . Hier soll nun ein Ausdruck für die Fehlerleistung gefunden werden, und zwar für den Spezialfall von weißen, gleichverteilten und unkorrelierten Rauschgrößen am Eingang; d.h. die Komponenten des Eingangsvektors

$$\mathbf{x}(k) = (x_1(k), x_2(k), \dots, x_r(k))^T \quad (4.41)$$

sollen Abtastwerte der von r unabhängigen Rauschquellen gelieferten Ausgangssignale sein.

Die Gleichverteilung der Signale x_0 wird hier so verstanden, daß bei Darstellung im Zweierkomplement der Bereich von 2^W diskreten Werten gleichmäßig ausgefüllt wird, der negative Extremwert $x = -1 = 1.0000 \dots$ ist also mit eingeschlossen:

$$-1 \leq x_0 \leq 1 - 2^{-(W-1)} \quad (4.42)$$

Dadurch verschwindet der Mittelwert der x_p nicht vollständig

$$\langle x_p \rangle = -2^{-w}, \quad (4.42b)$$

er fällt jedoch nur bei kleinen Eingangswortlängen ins Gewicht.

Wegen der Unabhängigkeit der r Rauschprozesse gilt die Gl. (4.42c), wenn der Wert 2^{-2w} hier, wie auch in den folgenden Abschnitten als Fehlergröße 2. Ordnung vernachlässigt wird.

$$\langle x_p(k) \cdot x_\sigma(k) \rangle = \begin{cases} 0 & \text{für } p \neq \sigma \\ \frac{1}{3} & \text{für } p = \sigma \end{cases} \quad (4.42c)$$

Aus den hier festgelegten Voraussetzungen für die Eingangsvariablen lassen sich statistische Eigenschaften der Bitmuster formulieren,

$$\langle \xi_{pi}(k) \rangle = \frac{1}{2} \quad p \in \{1, \dots, r\}, \quad i \in \{0, \dots, w-1\} \quad (4.43a)$$

$$\langle \xi_{pi}(k) \cdot \xi_{\sigma j}(k) \rangle = \begin{cases} \frac{1}{2} & \text{für } p = \sigma, i = j \\ \frac{1}{4} & \text{für } p \neq \sigma, i, j \in \{0, \dots, w-1\} \\ \frac{1}{4} & \text{für } i \neq j, p, \sigma \in \{1, \dots, r\} \end{cases} \quad (4.43b)$$

die im weiteren benötigt werden.

4.4.2 Berechnung der zeitlichen Mittelwerte

Für den Mittelwert von Δy_0 erhält man nach Gl. (4.32)

$$\langle \Delta y_0 \rangle = \sum_{p=1}^r \alpha_p \langle x_p(k) \rangle + Q_R \sum_{n=0}^{2^r-1} e_n^{(0)} \langle B(n, \mathbf{x}(k)) \rangle \quad (4.44a)$$

mit

$$\langle B(n, \mathbf{x}(k)) \rangle = - \sum_{p=1}^r (\xi_{p0}(k) \odot v_p) + \sum_{i=1}^{w-1} (2^{-i} \sum_{p=1}^r (\xi_{pi}(k) \odot v_p)) \quad (4.44b)$$

Da die Terme $(\xi_{\rho 1} \ominus v_{\rho})$ und $(\xi_{\sigma 1} \ominus v_{\sigma})$ in Gl. (4.44b) für $\rho \neq \sigma$ unabhängig sind, können Produkt- und Mittelungszeichen vertauscht werden. Außerdem folgt aus Gl. (4.43a) $\langle \xi_{\rho 1}(k) \ominus v_{\rho} \rangle = 1/2$, so daß sich für Gl. (4.44a) der einfache Ausdruck (4.45) ergibt.

$$\langle B(n, \mathbf{x}(k)) \rangle = -2^{-r} + 2^{-r} \sum_{i=1}^{w-1} 2^{-i} = -2^{-r} \cdot 2^{-(w-1)} \quad (4.45)$$

Mit Gl. (4.44a), (4.42b) und (4.45) erhält man den Mittelwert des resultierenden Fehlers:

$$\langle \Delta y_0(k) \rangle = -2^{-w} \left[2^{-(r-1)} Q_R \cdot \sum_{n=0}^{2^r-1} e_n^{(0)} + \sum_{\rho=1}^r \alpha_{\rho} \right] \quad (4.46)$$

Zur Berechnung von $\langle \Delta y_1 \rangle$ (siehe Gl. (4.39a)) kann der Ausdruck (4.46) verwendet werden:

$$\begin{aligned} \langle \Delta y_1(k) \rangle &= \sum_{\rho=1}^r \alpha_{\rho} \langle x_{\rho}(k) \rangle + Q_R \sum_{n=0}^{2^r-1} e_n^{(1)} \langle B(n, \mathbf{x}(k)) \rangle - \\ &- 2^{-(w-1)} Q_R e_{2^{r-1}}^{(1)} = \\ &= -2^{-w} \left[2^{-(r-1)} Q_R \cdot \sum_{n=0}^{2^r-1} e_n^{(1)} + \sum_{\rho=1}^r \alpha_{\rho} + 2 Q_R e_{2^{r-1}}^{(1)} \right] \quad (4.47) \end{aligned}$$

Nach Gl. (4.40b) verschwindet bei der modifizierten Form die Summe über alle Komponenten $e_n^{(1)}$, so daß sich Gl. (4.47) vereinfacht:

$$\langle \Delta y_1(k) \rangle = -2^{-w} (2 Q_R e_{2^{r-1}}^{(1)} + \sum_{\rho=1}^r \alpha_{\rho}) \quad (4.48)$$

4.4.3 Fehlerleistung bei Zweierkomplement-Darstellung der Eingangssignale

Zur Formulierung der Gesamtleistung von Δy_0 geht man von Gl. (4.32) aus:

$$\begin{aligned}
 N_{\Delta Y_O} &= \langle (\Delta Y_O(k))^2 \rangle = \\
 &= \sum_{\rho=1}^r \alpha_{\rho}^2 \langle x^2(k) \rangle + 2Q_R \cdot \sum_{\rho=1}^r \sum_{n=0}^{2^r-1} \alpha_{\rho} e_n^{(0)} \langle x_{\rho}(k) \cdot B(n, \mathbf{x}(k)) \rangle + \\
 &+ Q_R^2 \sum_{n=0}^{2^r-1} \sum_{m=0}^{2^r-1} e_n^{(0)} e_m^{(0)} \langle B(n, \mathbf{x}(k)) \cdot B(m, \mathbf{x}(k)) \rangle \quad (4.49)
 \end{aligned}$$

Unter Berücksichtigung der in Abschnitt 4.4.1 angeführten statistischen Eigenschaften der Eingangssignale (4.42), (4.43) findet man nach einigem Rechenaufwand (siehe Anhang 3 und 4) die folgenden Ausdrücke für die verschiedenen Terme in Gl. (4.49):

$$\langle B(n, \mathbf{x}(k)) \cdot B(m, \mathbf{x}(k)) \rangle = 4/3 \cdot 2^{-r} (\delta_{mn} - 2^{-r}) \quad (4.50)$$

mit dem Kronecker-Symbol

$$\delta_{nm} = \begin{cases} 1 & \text{für } n = m \\ 0 & \text{für } n \neq m \end{cases}$$

$$\langle x_{\rho}(k) \cdot B(n, \mathbf{x}(k)) \rangle = \begin{cases} 1/3 \cdot 2^{-r+1} & \text{für } v_{\rho} = 1 \\ -1/3 \cdot 2^{-r+1} & \text{für } v_{\rho} = 0 \end{cases} \quad (4.51)$$

Für (4.51) läßt sich eine geschlossene Darstellung finden, bei der nicht mehr die Bits v_{ρ} der Binärdarstellung von n benötigt werden, sondern der Index n selbst benutzt wird: Aus Gl. (4.28b) ergibt sich die Beziehung

$$\frac{n}{2^{\rho-1}} = \sum_{\sigma=1}^r v_{\sigma} 2^{\sigma-\rho}, \quad (4.52a)$$

und man erhält für die größte ganze Zahl, die kleiner oder gleich $\frac{n}{2^{\rho-1}}$ ist

$$\left\lfloor \frac{n}{2^{\rho-1}} \right\rfloor = \sum_{\sigma=\rho}^r v_{\sigma} \cdot 2^{\sigma-\rho} = \begin{cases} \text{gerade für } v_{\rho} = 0 \\ \text{ungerade für } v_{\rho} = 1 \end{cases} \quad (4.52b)$$

Es folgt damit aus Gl. (4.51)

$$\langle x_{\rho}(k) \cdot B(n, \mathbf{x}(k)) \rangle = -(-1)^{\left\lfloor \frac{2n}{2^{\rho}} \right\rfloor} \cdot 1/3 \cdot 2^{-r+1}. \quad (4.53)$$

Die Gleichungen (4.42), (4.43), (4.49), (4.50) und (4.53) führen schließlich zu einem geschlossen analytischen Ausdruck für die Fehlerleistung, der zur Beschreibung des individuellen zu untersuchenden Systems nur die a-priori-Koeffizientenfehler und den Fehlervektor enthält.

$$N_{\Delta Y_0} = \frac{1}{3} \sum_{\rho=1}^r \left(\alpha_{\rho} 2^{-r+2} \cdot Q_R \cdot \alpha_{\rho} \sum_{n=0}^{2^r-1} (-1)^{\left\lfloor \frac{2n}{2^{\rho}} \right\rfloor} e_n^{(0)} \right) + \\ + \frac{4}{3} Q_R 2 \cdot 2^{-r} \left(\sum_{n=0}^{2^r-1} (e_n^{(0)})^2 2^{-r} \left(\sum_{n=0}^{2^r-1} e_n^{(0)} \right)^2 \right) \quad (4.54)$$

In Abschnitt 4.4.2 wurde gezeigt, daß der Fehler ΔY_0 wegen der endlichen Wortlänge w der Eingangsvariablen einen Gleichanteil besitzt (4.46), so daß seine Varianz nicht mit der Gesamtleistung übereinstimmt:

$$\sigma_{\Delta Y_0}^2 = \langle (\Delta Y_0)^2 \rangle - \langle \Delta Y_0 \rangle^2 = \\ = N_{\Delta Y_0} - 2^{-2w} (2^{-(r-1)} Q_R \sum_{n=0}^{2^r-1} e_n^{(0)} + \sum_{\rho=1}^r \alpha_{\rho})^2 \quad (4.55)$$

Der Unterschied ist allerdings nur bei extrem kleinen Eingangswortlängen von Bedeutung; in üblichen Anwendungen (z.B. $w \geq 8$ bit) ist $\langle \Delta Y_0 \rangle^2$ zu vernachlässigen.

4.4.4 Fehlerleistung beim modifizierten Zweierkomplement

Subtrahiert man von dem in Gl. (4.39) angegebenen Ausdruck für Δy_1 den konstanten Wert

$$A = - 2^{-(w-1)} Q_R \cdot e_{2^r-1}^{(1)}, \quad (4.56a)$$

so erhält man die gleiche Form wie in Gl. (4.32), also die Fehlergröße Δy_0 .

D.h. für den quadratischen Mittelwert $\langle (\Delta y_1 - A)^2 \rangle$ ist formal der Ausdruck für die Gesamtleistung von Δy_0 (Gl. (4.54)) zu setzen,

wobei $e_n^{(0)}$ durch $e_n^{(1)}$ zu ersetzen ist.

$$\begin{aligned} \langle (\Delta Y_1 - A)^2 \rangle &= \frac{1}{3} \sum_{\rho=1}^r \left(\alpha_{\rho}^2 \cdot 2^{-r+2} Q_R \alpha_{\rho} \sum_{n=0}^{2^r-1} (-1)^{\left[\frac{2n}{2^{\rho}} \right]} e_n^{(1)} \right) + \\ &+ \frac{4}{3} Q_R^2 \cdot 2^{-r} \left(\sum_{n=0}^{2^r-1} (e_n^{(1)})^2 \cdot 2^{-r} \left(\sum_{n=0}^{2^r-1} e_n^{(1)} \right)^2 \right) \quad (4.57) \end{aligned}$$

Daraus ist die Gesamtleistung für die modifizierte Form zu bestimmen

$$N_{\Delta Y_1} = \langle (\Delta Y_1)^2 \rangle = \langle (\Delta Y_1 - A)^2 \rangle + 2A \langle \Delta Y_1 \rangle - A^2, \quad (4.58a)$$

indem der Mittelwert $\langle \Delta Y_1 \rangle$ nach Gl. (4.48) eingesetzt wird.

Man erhält schließlich unter Berücksichtigung von $\sum_{n=0}^{2^r-1} e_n^{(1)} = 0$:

$$\begin{aligned} N_{\Delta Y_1} &= \frac{1}{3} \sum_{\rho=1}^r \left(\alpha_{\rho}^2 \cdot 2^{-r+2} Q_R \alpha_{\rho} \sum_{n=0}^{2^r-1} (-1)^{\left[\frac{2n}{2^{\rho}} \right]} e_n^{(1)} \right) + \\ &+ \frac{4}{3} Q_R^2 \cdot 2^{-r} \cdot \sum_{n=0}^{2^r-1} (e_n^{(1)})^2 + 2^{-2(w-1)} \left(Q_R^2 (e_{\frac{1}{2}^{r-1}}^{(1)})^2 + \right. \\ &\left. + Q_R e_{\frac{1}{2}^{r-1}}^{(1)} \cdot \sum_{\rho=1}^r \alpha_{\rho} \right) \quad (4.58b) \end{aligned}$$

In den meisten Anwendungsfällen ist die Eingangswortlänge w so groß, daß der letzte Term in (4.58) zu vernachlässigen ist und man vereinfachend schreiben kann:

$$\begin{aligned} N_{\Delta Y_1} &\approx \sigma_{\Delta Y_1}^2 \approx \frac{1}{3} \sum_{\rho=1}^r \left(\alpha_{\rho}^2 \cdot 2^{-r+2} Q_R \alpha_{\rho} \sum_{n=0}^{2^r-1} (-1)^{\left[\frac{2n}{2^{\rho}} \right]} e_n^{(1)} \right) + \\ &+ \frac{4}{3} Q_R^2 \cdot 2^{-r} \sum_{n=0}^{2^r-1} (e_n^{(1)})^2 \quad (4.59) \end{aligned}$$

4.5 Definition der Rest-Koeffizientenfehler

In Abschnitt 4.1 wurde erläutert, daß die linearen Verzerrungen infolge individueller Koeffizientenfehler bei der Verteilten Arithmetik zurückgehen, und zwar um so stärker, je mehr Koeffizienten beteiligt sind. Um diese Eigenschaft genauer zu untersuchen, muß zunächst für das nichtlineare System "Vektor-Multiplizierer in Verteilter Arithmetik" ein Verfahren zur Bestimmung der linear mit den Eingangssignalen korrelierten Anteile im Gesamtfehler entwickelt werden.

Aus der Regelungstechnik ist eine Methode bekannt, bei der nicht-lineare Systeme, die mit stochastischen Signalen erregt werden, mit Hilfe eines linearen Ersatzmodells beschrieben werden ([26, 27]). Dazu wird eine äquivalente, lineare Verstärkung $K(\sigma_x)$ definiert, die ein Maß für den im Ausgangssignal $y(t)$ enthaltenen linear mit dem Eingangssignal $x(t)$ korrelierten Anteil darstellt. Wegen der Nichtlinearität des betrachteten Systems ist die äquivalente Verstärkung von der Leistung des Eingangssignals abhängig.

Aus dem in Bild 4.6 wiedergegebenen Prinzip des linearen Ersatzsystems ergibt sich die Definition einer Extremalwert-Aufgabe, die zur Berechnung der äquivalenten Verstärkung K führt.

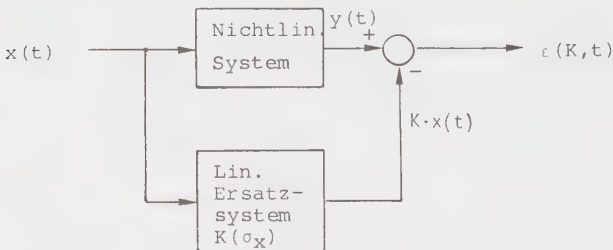


Bild 4.6: Nichtlineares System mit linearem Ersatzsystem

Man bestimmt dabei die Verstärkung K derart, daß die Leistung der Abweichung $\epsilon(k)$ zwischen nichtlinearem und linearem System minimal wird:

$$\begin{aligned} \langle \varepsilon^2(K, t) \rangle &= \langle (y(t) - Kx(t))^2 \rangle = \langle y^2(t) \rangle - 2K \langle x(t) \cdot y(t) \rangle + \\ &+ K^2 \langle x^2(t) \rangle \stackrel{!}{=} \text{Min} \end{aligned} \quad (4.60)$$

Gl. (4.60) ist unmittelbar zu lösen:

$$\frac{\partial \langle \varepsilon^2(K, t) \rangle}{\partial K} = -2 \langle x(t) \cdot y(t) \rangle + 2K \langle x^2(t) \rangle \stackrel{!}{=} 0 \quad (4.61a)$$

$$K = \frac{\langle x(t) \cdot y(t) \rangle}{\langle x^2(t) \rangle} \quad (4.61b)$$

Wird dieser Wert für das lineare Ersatzsystem gesetzt, so sind $x(t)$ und $\varepsilon(t)$ nicht mehr linear korreliert:

$$\langle x(t) \cdot \varepsilon(K, t) \rangle = \langle x(t) \cdot y(t) \rangle - \frac{\langle x(t) \cdot y(t) \rangle}{\langle x^2(t) \rangle} \cdot \langle x^2(t) \rangle = 0 \quad (4.61c)$$

Bei der Untersuchung des Multipliziererrohrs in digitalen Filtern wurde das hier beschriebene Verfahren bereits früher verwendet [28,29]. Um es auf die Fehleranalyse der Verteilten Arithmetik anzuwenden, muß es auf Systeme mit mehreren Eingängen erweitert werden (siehe Bild 4.7).

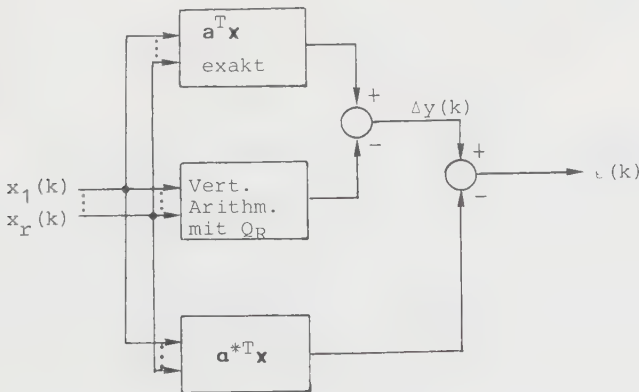


Bild 4.7: Definition der äquivalenten Verstärkung für ein System mit r Eingängen.

Der Vektor $\mathbf{a}^* = (\alpha_1^*, \dots, \alpha_r^*)^T$ gibt die Proportionalitäts-Faktoren der im Fehler Δy enthaltenen linear mit den Eingangssignalen korrelierten Anteile an. Sie sollen im folgenden als "Rest-Koeffizientenfehler" bezeichnet werden, im Gegensatz zu den a-priori-Koeffizientenfehlern der Konzentrierten Arithmetik.

Zur Berechnung der Rest-Koeffizientenfehler geht man von dem Ansatz (4.62) aus

$$\Delta y = \alpha_1^* x_1 + \alpha_2^* x_2 + \dots + \alpha_r^* x_r + \varepsilon. \quad (4.62)$$

Daraus erhält man unter der Bedingung, daß die Eingangsvariablen untereinander und mit der Größe ε nicht korreliert sind

$$\langle x_\rho \cdot x_\sigma \rangle = \begin{cases} 0 & \text{für } \rho \neq \sigma \\ \frac{1}{3} & \text{für } \rho = \sigma \end{cases} \quad (4.63a)$$

$$\langle x_\rho \cdot \varepsilon \rangle = 0 \quad \forall \rho \in \{1, \dots, r\} \quad (4.63b)$$

die gesuchte Formulierung der Rest-Koeffizientenfehler:

$$\langle \Delta y \cdot x_\rho \rangle = \sum_{\sigma=1}^r \langle x_\rho \cdot x_\sigma \rangle \cdot \alpha_\sigma^* + \langle x_\rho \cdot \varepsilon \rangle = \frac{1}{3} \alpha_\rho^* \quad (4.64a)$$

$$\alpha_\rho^* = 3 \langle \Delta y x_\rho \rangle ; \rho \in \{1, 2, \dots, r\} . \quad (4.64b)$$

4.6 Analytische Beschreibung der Rest-Koeffizientenfehler

Für die Berechnung der Rest-Koeffizientenfehler nach Gl. (4.64b) werden die Beziehungen (4.32) und (4.39) benutzt. Man erhält für Zweierkomplement und modifizierte Form die Ausdrücke

$$\frac{\alpha_\rho^*(0)}{3} = \alpha_\rho \langle x_\rho^2(k) \rangle + Q_R \sum_{n=0}^{2^r-1} (e_n^{(0)} \langle x_\rho(k) \cdot B(n, \mathbf{x}(k)) \rangle) \quad (4.65a)$$

$$\frac{\alpha_\rho^*(1)}{3} = \alpha_\rho \langle x_\rho^2(k) \rangle + Q_R \sum_{n=0}^{2^r-1} (e_n^{(1)} \langle x_\rho(k) \cdot B(n, \mathbf{x}(k)) \rangle) -$$

$$- 2^{-(w-1)} Q_R e_{2^{r-1}}^{(1)} \cdot \langle x_p(k) \rangle \quad (4.65b)$$

In Gl. (4.65b) ist der Term

$$- 2^{-w+1} Q_R e_{2^{r-1}}^{(1)} \cdot \langle x_p(k) \rangle = 2^{-2w+1} Q_R e_{2^{r-1}}^{(1)}$$

für hinreichend große Eingangs-Wortlängen w zu vernachlässigen, so daß sich für $\alpha_p^{*(0)}$ und $\alpha_p^{*(1)}$ formal die gleichen Ausdrücke ergeben.

Der Term $\langle x_p(k) \cdot B(n, x(k)) \rangle$ wurde in Abschnitt 4.43 bereits bestimmt, so daß man mit Gl. (4.51) und (4.65a,b) die analytische Formulierung der Rest-Koeffizientenfehler erhält:

$$\alpha_p^{*(0,1)} = \alpha_p - Q_R 2^{-r+1} \sum_{n=0}^{2^r-1} (-1)^{\left\lfloor \frac{2n}{2^p} \right\rfloor} e_n^{(0,1)} \quad (4.66)$$

Damit läßt sich nun vom Gesamtfehler Δy ein Term abspalten, der nicht mehr mit den Eingangssignalen korreliert ist. Seine Leistung ist unter Verwendung von Gl. (4.32) und (4.66) anzugeben (vgl. Anhang 5):

$$N_E = \langle (\Delta y - \sum_{p=1}^r \alpha_p^* x_p)^2 \rangle \quad (4.67)$$

$$N_{E_0} = \frac{4}{3} Q_R^2 2^{-r} \left(\sum_{n=0}^{2^r-1} (e_n^{(0)})^2 - 2^{-r} \left(\sum_{n=0}^{2^r-1} e_n^{(0)} \right)^2 + \right. \\ \left. + \sum_{p=1}^r \left(\sum_{n=0}^{2^r-1} (-1)^{\left\lfloor \frac{2n}{2^p} \right\rfloor} e_n^{(0)} \right)^2 \right) \quad (4.68a)$$

$$N_{E_1} = \frac{4}{3} Q_R^2 2^{-r} \left(\sum_{n=0}^{2^r-1} (e_n^{(1)})^2 - 2^{-r} \sum_{p=1}^r \left(\sum_{n=0}^{2^r-1} (-1)^{\left\lfloor \frac{2n}{2^p} \right\rfloor} e_n^{(1)} \right)^2 \right) \quad (4.68b)$$

Zusammenfassend ist zu der bis hierher dargelegten Fehleranalyse das folgende festzuhalten:

Es wird für gleichverteilte, spektral weiße und untereinander nicht korrelierte Rauschsignale am Eingang ein geschlossener analytischer Ausdruck für die Fehler-Gesamtleistung angegeben (Gl. 4.54, 4.59).

Dieser Fehler ist aufzuspalten in linear mit den Eingangssignalen korrelierte Terme und einen unkorrelierten Anteil. Das Fehlerverhalten ist damit zu beschreiben durch lineare Verzerrungen, die durch die in Gl. (4.66) angegebenen Rest-Koeffizientenfehler bestimmt sind, und eine unabhängige Rauschquelle am Ausgang mit der in Gl. (4.68a,b) angegebenen Leistung. Korrelationen höherer Ordnung bleiben bei diesem Modell unberücksichtigt. Wenn also im folgenden die Rede ist von "unkorrelierten" Fehleranteilen, so ist dieses stets im Sinne dieses Modells zu verstehen: D.h. dieser Anteil ist n i c h t l i n e a r k o r r e l i e r t mit den Eingangssignalen.

4.7 Untersuchung einzelner Beispiele

4.7.1 Überprüfung der Fehleranalyse durch Simulation

Die in den vorausgegangenen Abschnitten hergeleiteten Beziehungen sollen durch die Simulation einiger Beispiele überprüft werden. Dazu muß das in Kap. 4.2.2 verwendete Simulations-Blockdiagramm (Bild 4.2) für die Bestimmung der Rest-Koeffizientenfehler erweitert werden.

Die Werte α_{ρ}^* werden hier entsprechend ihrer Definition (Gl. 4.64) berechnet. Die Leistung des unkorrelierten Fehlers N_{ϵ} kann im Anschluß an den Simulationslauf wegen der Unabhängigkeit der Eingangssignale x_{ρ} nach Gl. (4.69) bestimmt werden

$$N_{\epsilon} = N_{\Delta y} - \frac{1}{3} \sum_{\rho=1}^K (\alpha_{\rho}^*)^2 . \quad (4.69)$$

Zur Simulation werden die bereits in Abschnitt 4.2.2 benutzten Beispiele herangezogen. Für die Mittelwertberechnung wird eine Stichprobenmenge von $L = 50\,000$ verwendet, die Quantisierungsstufe der Tafel wird mit $Q_R = 2^{-7}$ festgelegt.

Die anschließende Aufstellung zeigt die analytisch bestimmten Werte für die gesamte Fehlerleistung (nach Gl. (4.54), (4.59)), Rest-Koeffizientenfehler (Gl. 4.66) und Leistung des unkorrelier-

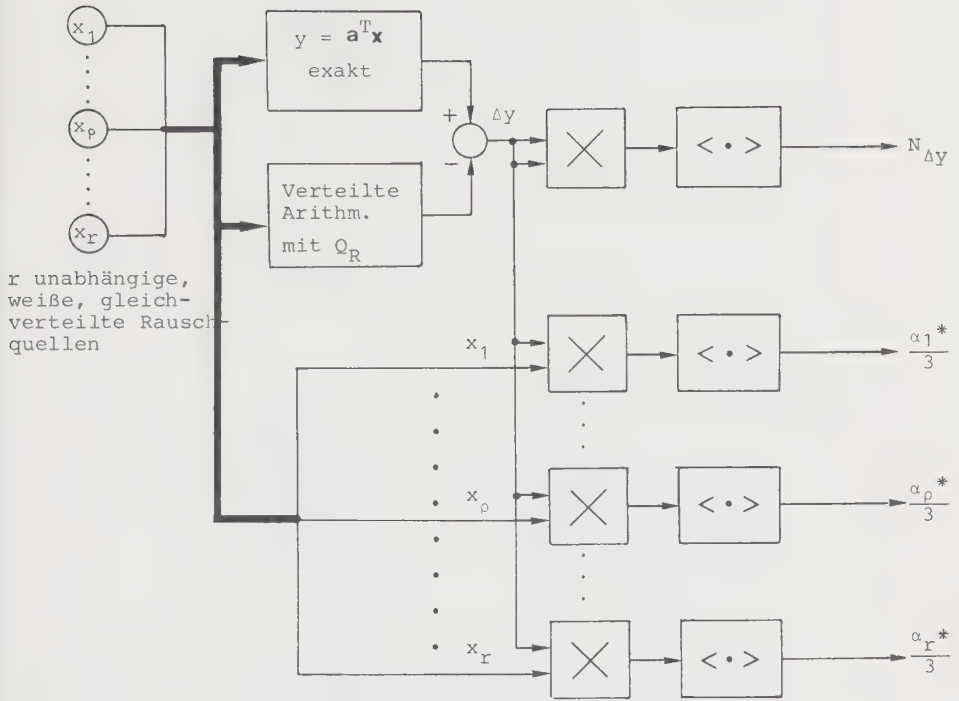


Bild 4.8: Blockdiagramm zur Simulation

ten Fehlers (Gl. (4.68a,b)) im Vergleich zu den gemäß Bild 4.8 gewonnenen Simulationsergebnissen.

Beispiel 1 ($r = 2$)

$$a_1 = 0.2178328$$

$$\alpha_1 = -0.1174 Q_R$$

$$a_2 = -0.3159031$$

$$\alpha_2 = -0.4356 Q_R$$

Zweierkomplement:

	Analyse	Simulation
$N_{\Delta y_0}$	$0.13351 \ Q_R^2$	$0.1331 \ Q_R^2$
N_{ϵ_0}	$0.08333 \ Q_R^2$	$0.0827 \ Q_R^2$
$\alpha_1^{*(0)}$	$0.3826 \ Q_R$	$0.3833 \ Q_R$
$\alpha_2^{*(0)}$	$0.0644 \ Q_R$	$0.0650 \ Q_R$

Modifiziertes Zweierkomplement

	Analyse	Simulation
$N_{\Delta y_1}$	$0.05018 \ Q_R^2$	$0.0504 \ Q_R^2$
N_{ϵ_1}	0.0000	$0.00004 \ Q_R^2$
$\alpha_1^{*(1)}$	$0.3826 \ Q_R$	$0.3835 \ Q_R$
$\alpha_2^{*(1)}$	$0.0644 \ Q_R$	$0.0651 \ Q_R$

Beispiel_2 (r = 2)

$$a_1 = 0.4941047$$

$$\alpha_1 = 0.2454 \ Q_R$$

$$a_2 = 0.1852891$$

$$\alpha_2 = -0.2830 \ Q_R$$

Zweierkomplement

	Analyse	Simulation
$N_{\Delta y_0}$	$0.04677 \ Q_R^2$	$0.04669 \ Q_R^2$
N_{ϵ_0}	0.0000	$0.00003 \ Q_R^2$
$\alpha_1^{*(0)}$	$0.2454 \ Q_R$	$0.2451 \ Q_R$
$\alpha_2^{*(0)}$	$-0.2830 \ Q_R$	$-0.2830 \ Q_R$

Modifiziertes Zweierkomplement

	Analyse	Simulation
$N_{\Delta Y_1}$	$0.03730 Q_R^2$	$0.03735 Q_R^2$
$N_{\varepsilon 1}$	0.0000	$0.00025 Q_R^2$
$\alpha_1^{*(1)}$	$-0.2546 Q_R$	$-0.2526 Q_R$
$\alpha_2^{*(1)}$	$0.2170 Q_R$	$0.2179 Q_R$

Beispiel_3 (r = 8):

$a_1 = 0.251531$	$\alpha_1 = 0.196 Q_R$
$a_2 = 0.218546$	$\alpha_2 = -0.026 Q_R$
$a_3 = -0.405625$	$\alpha_3 = 0.080 Q_R$
$a_4 = 0.849391$	$\alpha_4 = -0.278 Q_R$
$a_5 = -0.260927$	$\alpha_5 = -0.399 Q_R$
$a_6 = 0.282398$	$\alpha_6 = 0.147 Q_R$
$a_7 = 0.081344$	$\alpha_7 = 0.412 Q_R$
$a_8 = 0.155047$	$\alpha_8 = 0.154 Q_R$

Zweierkomplement

	Analyse	Simulation
$N_{\Delta Y_0}$	$0.10789 Q_R^2$	$0.1079 Q_R^2$
$N_{\varepsilon 0}$	$0.10610 Q_R^2$	$0.1061 Q_R^2$
$\alpha_1^{*(0)}$	$0.0160 Q_R$	$0.0180 Q_R$
$\alpha_2^{*(0)}$	$-0.0031 Q_R$	$-0.0028 Q_R$
$\alpha_3^{*(0)}$	$-0.0211 Q_R$	$-0.0212 Q_R$

	Analyse	Simulation
$\alpha_4^{*(0)}$	-0.0201 Q_R	-0.0187 Q_R
$\alpha_5^{*(0)}$	-0.0410 Q_R	-0.0426 Q_R
$\alpha_6^{*(0)}$	-0.0019 Q_R	-0.0012 Q_R
$\alpha_7^{*(0)}$	0.0450 Q_R	0.0444 Q_R
$\alpha_8^{*(0)}$	0.0060 Q_R	0.0063 Q_R

Modifiziertes Zweierkomplement

	Analyse	Simulation
$N_{\Delta Y_1}$	0.028193 Q_R^2	0.0282 Q_R^2
$N_{\varepsilon 1}$	0.027850 Q_R^2	0.0277 Q_R^2
$\alpha_1^{*(1)}$	-0.0232 Q_R	-0.0240 Q_R
$\alpha_2^{*(1)}$	0.0047 Q_R	0.0041 Q_R
$\alpha_3^{*(1)}$	-0.0132 Q_R	-0.0127 Q_R
$\alpha_4^{*(1)}$	-0.0124 Q_R	-0.0117 Q_R
$\alpha_5^{*(1)}$	-0.0019 Q_R	-0.0019 Q_R
$\alpha_6^{*(1)}$	-0.0098 Q_R	-0.0102 Q_R
$\alpha_7^{*(1)}$	0.0059 Q_R	0.0063 Q_R
$\alpha_8^{*(1)}$	-0.0018 Q_R	-0.0021 Q_R

Die gute Übereinstimmung von Simulation und Analyse bestätigt die abgeleiteten Formeln.

4.7.2 Einige Eigenschaften im Fehlerverhalten der Verteilten Arithmetik

Vergleicht man die im letzten Abschnitt gefundenen Ergebnisse mit denen von Abschnitt 4.2.3, so erweisen sich die beiden Darstellungen der Gesamtfehlerleistung (Gl. 4.21 a,b und Gl. 4.54, 4.59) als äquivalent. Der Vorteil der Formulierung nach (4.54, 4.59) liegt jedoch in der geschlossenen Darstellung auf der Basis des Fehlervektors $\mathbf{e}$, der in überschaubarer Form die Informationen über die Fehler-Charakteristik des betrachteten individuellen Koeffizientensatzes enthält. Er liefert dann auch eine Aussage darüber, inwieweit der nach Gl. (4.21a,b) und Gl. (4.54, 4.59) gleichermaßen angebbare Gesamtfehler aufgrund linearer Verzerrungen zustande kommt, oder durch eine unabhängige Rauschquelle zu beschreiben ist. Betrachtet man nämlich die Rest-Koeffizientenfehler-Analyse der beiden Fälle für $r = 2$ bei Zweierkomplement-Codierung, so wird ein prinzipieller Unterschied deutlich, der bei der Analyse in Abschnitt 4.2.3 unbemerkt blieb: Während in Beispiel 1 ein nicht korrelierter Anteil im Gesamtfehler vorhanden ist, verschwindet dieser in Beispiel 2. Im Fall 2 sind die Rest-Koeffizientenfehler identisch mit den a-priori-Koeffizientenfehlern. Bei der modifizierten Form verschwindet die Leistung N_{ε_1} bei beiden Beispielen; die Koeffizientenfehler sind in beiden Fällen verschieden von den a-priori-Koeffizientenfehlern.

Anhand der oben abgeleiteten Beziehungen läßt sich dieses Fehlerverhalten erläutern.

Aus der Definition der Fehlervektor-Komponenten (Gl. 4.28a) ergibt sich für Zweierkomplement-Codierung im Fall $r = 2$

$$\begin{aligned} e_0^{(0)} &= e_1^{(0)} = e_2^{(0)} = 0 \\ e_3^{(0)} &\in \{0, +1, -1\} , \end{aligned} \quad (4.70a)$$

so daß man für die Leistung des unkorrelierten Fehlers mit (4.68a) erhält:

$$N_{\epsilon_0} \Big|_{r=2} = \frac{1}{12} Q_R^2 \cdot (e_3^{(0)})^2 =$$

$$= \begin{cases} 0 & \text{für } e_3^{(0)} = 0 \quad (\text{vgl. Beispiel 2}) \\ Q_R^2/12 & \text{für } e_3^{(0)} = \pm 1 \quad (\text{vgl. Beispiel 1}) \end{cases} \quad (4.70b)$$

Die Leistung des unkorrelierten Fehleranteils kann also für $r = 2$ überhaupt nur 2 Werte annehmen. Die jeweils zugehörigen Rest-Koeffizientenfehler bestimmt man aus Gl. (4.66)

$$\alpha_\rho^{*(0)} \Big|_{r=2} = \alpha_\rho + \frac{Q_R}{2} e_3^{(0)} = \begin{cases} \alpha_\rho & \text{für } e_3^{(0)} = 0 \quad (\text{vgl. Beispiel 2}) \\ \alpha_\rho \pm Q_R/2 & \text{für } e_3^{(0)} = \pm 1 \quad (\text{vgl. Beispiel 1}) \end{cases} \quad (4.70c)$$

Bei der modifizierten Form gilt für den Fehlervektor nach Definition Gl. (4.38b)

$$e_0^{(1)} = -e_3^{(1)} ; \quad e_1^{(1)} = -e_2^{(1)} ; \quad (4.71a)$$

$$e_n \in \{0, +\frac{1}{2}, -\frac{1}{2}\} \quad \text{für zwei Koeffizienten .}$$

Aus Gl. (4.68b) erhält man damit

$$N_{\epsilon_1} \Big|_{r=2} = \frac{Q_R^2}{3} (2e_0^2 + 2e_1^2 - \frac{1}{4} ((2e_0 - 2e_1)^2 + (2e_0 + 2e_1)^2)) =$$

$$= \frac{Q_R^2}{3} (2e_0^2 + 2e_1^2 - (e_0^2 - 2e_0e_1 + e_1^2 + e_0^2 + 2e_0e_1 + e_1^2)) = 0 \quad (4.71b)$$

Die endliche Tafelwortlänge führt also bei der modifizierten Form für den Fall $r = 2$ zu einem Fehler, der ausschließlich linear mit den Eingangssignalen korreliert ist. Interessant ist in diesem Zusammenhang die Feststellung, daß die Realisierung dieses Sonderfalles die gleiche Kapazität für den Koeffizientenspeicher erfordert wie die Konzentrierte Arithmetik [7], in beiden Fällen entstehen durch endliche Koeffizientenspeicher-Wortlängen nur lineare Verzerrungen. Allerdings zeigen die Beispiele

1 und 2, daß sich bei der Verteilten Arithmetik für die Rest-Koeffizientenfehler nicht notwendig die a-priori-Koeffizientenfehler ergeben. Die Anwendung der Gl. (4.66) im Fall $r = 2$ macht dies deutlich:

$$\begin{aligned}\alpha_1^{*(1)} &= \alpha_1 - Q_R (e_0 - e_1) \\ \alpha_2^{*(1)} &= \alpha_2 - Q_R (e_0 + e_1)\end{aligned}\tag{4.71c}$$

Der Wertevorrat für die Fehlervektor-Komponenten ist in (4.71a) angegeben. Dabei sind also nicht alle beliebigen Kombinationen für e_0 und e_1 möglich, wie aus der Definitionsgleichung (4.38b) unmittelbar ersichtlich ist.

In Gleichung (4.71d) sind die erforderlichen Fallunterscheidungen berücksichtigt.

$$\begin{aligned}\left. \begin{aligned}\alpha_1^{*(1)} &= \alpha_1 + Q_R/2 \\ \alpha_2^{*(1)} &= \alpha_2 + Q_R/2\end{aligned} \right\} & \text{für } e_0 = \pm \frac{1}{2}, e_1 = 0 \text{ (vgl. Beisp. 1)} \\ \left. \begin{aligned}\alpha_1^{*(1)} &= \alpha_1 \pm Q_R/2 \\ \alpha_2^{*(1)} &= \alpha_2 + Q_R/2\end{aligned} \right\} & \text{für } e_0 = 0, e_1 = \pm \frac{1}{2} \text{ (vgl. Beisp. 2)} \\ \left. \begin{aligned}\alpha_1^{*(1)} &= \alpha_1 \\ \alpha_2^{*(1)} &= \alpha_2\end{aligned} \right\} & \text{für } e_0 = 0, e_1 = 0\end{aligned}\tag{4.71d}$$

Der Fall rein linearer Verzerrungen kann selbstverständlich auch bei Werten $r > 2$ auftreten; Beispiel 4*) zeigt einen entsprechenden Koeffizientensatz für $r = 4$. Mit steigender Koeffizientenanzahl wird jedoch die Wahrscheinlichkeit dafür immer geringer, wenn die a-priori-Koeffizientenfehler als zufällig verteilt angesehen werden. So können Systeme mit 8 Koeffizienten

*) Weitere Beispiele 4, 5, 6 sind am Schluß dieses Kapitels zusammengestellt.

in der Regel bereits mit guter Näherung allein durch eine unabhängige Rauschquelle am Ausgang beschrieben werden, während lineare Verzerrungen vernachlässigt werden: In Beispiel 3 macht die Leistung des unkorrelierten Fehleranteils 98,3 % (Zweierkomplement) bzw. 98,8 % (Modifizierte Form) der Gesamtleistung aus, die Rest-Koeffizientenfehler sind stark reduziert.

Auf eine Besonderheit bezüglich der Rest-Koeffizientenfehler soll abschließend hingewiesen werden: Auch bei größeren Werten r kann sich bei bestimmten Koeffizienten-Konstellationen der Fall ergeben, daß ein oder mehrere Rest-Koeffizientenfehler den gleichen Wert besitzen wie die zugehörigen a-priori-Koeffizientenfehler, während die anderen davon abweichen. Beispiel 5 zeigt den Fall eines solchen Koeffizientensatzes mit $\alpha_3^{*(0)} = \alpha_3$, $\alpha_4^{*(0)} = \alpha_4$. Es folgt hieraus allerdings noch nicht, daß der entstehende unkorrelierte Rauschanteil durch die α_3 und α_4 nicht beeinflusst wird.

Dazu seien für die Zweierkomplement-Codierung folgende Überlegungen angeführt:

Für den Fall, daß einzelne Rest-Koeffizientenfehler mit den a-priori Koeffizientenfehlern übereinstimmen, lassen sich eindeutige Bedingungen für die Fehlervektor-Komponenten angeben, z.B. für $\rho = r$ (was keine Beschränkung der Allgemeinheit bedeutet):

$$\alpha_r^{*(0)} = \alpha_r - Q_R \cdot 2^{-r+1} \sum_{n=0}^{2^r-1} (-1)^{\left\lceil \frac{2n}{2^r} \right\rceil} e_n^{(0)} \stackrel{!}{=} \alpha_r \quad (4.72a)$$

$$\sum_{n=0}^{2^{r-1}-1} e_n^{(0)} = \sum_{n=2^{r-1}}^{2^r-1} e_n^{(0)} \quad (4.72b)$$

In Gl. (4.72b) ist die linke Seite unabhängig von α_r , da aus $n < 2^{r-1}$ folgt $v_r \equiv 0$ (vgl. hierzu Gl. (4.28)). Auf der rechten Seite ($n \geq 2^{r-1}$) geht bei der Berechnung der $e_n^{(0)}$ jeweils α_r mit ein. Wird hierbei i n f o l g e α_r bei irgendeiner Kombination die Quantisierungsstufe übersprungen, so muß das (wegen 4.72b) bei einer anderen Kombination ebenfalls i n f o l g e α_r wieder rückgängig gemacht werden, d.h. eine

Quantisierungsstufe in entgegengesetzter Richtung überschritten werden. Das ist aber deshalb nicht möglich, weil der Einfluß von α_r bei jeder Kombination in die gleiche Richtung geht. Es folgt daraus, daß der Fehlervektor unter der Bedingung (4.72a) von α_r nicht beeinflußt werden kann, und damit auch nicht der Ausdruck N_{ϵ_O} und die anderen Rest-Koeffizientenfehler $\alpha_{\rho}^{*(O)}$ ($\rho \neq r$). In Beispiel 6 wird das deutlich: Gegenüber Beispiel 5 sind hier a_4 und a_5 Null gesetzt; am übrigen Fehlerverhalten ändert sich nichts.

Damit ist gezeigt, daß besondere Fälle existieren, in denen einzelne Eingangsvariablen sich bezüglich des Quantisierungsfehlers des Koeffizientenspeichers verhalten, wie wenn das System in Konzentrierter Arithmetik realisiert wäre, während das System für alle anderen Eingangssignale nichtlinearen Charakter hat.

Zusammenstellung einiger weiterer Beispiele (Q_R jeweils 2^{-7})

Beispiel_4:

$r = 4$	$a_1 = 0.24817188$	$\alpha_1 = -0.234 Q_R$
	$a_2 = -0.35846875$	$\alpha_2 = 0.116 Q_R$
	$a_3 = 0.17246875$	$\alpha_3 = 0.076 Q_R$
	$a_4 = 0.13389844$	$\alpha_4 = 0.139 Q_R$
	$N_{\Delta y_O} = 0.0311 Q_R^2$	$\alpha_1^{*(O)} = -0.234 Q_R$
		$\alpha_2^{*(O)} = 0.116 Q_R$
		$\alpha_3^{*(O)} = 0.076 Q_R$
	$N_{\epsilon_O} = 0.000$	$\alpha_4^{*(O)} = 0.139 Q_R$

Beispiel_5:

$r = 6$	$a_1 = 0.52855938$	$\alpha_1 = -0.3444 Q_R$
	$a_2 = -0.20002969$	$\alpha_2 = 0.3962 Q_R$
	$a_3 = 0.65689141$	$\alpha_3 = 0.0821 Q_R$

$$a_4 = 0.35950313$$

$$\alpha_4 = -0.0164 Q_R$$

$$a_5 = -0.12233516$$

$$\alpha_5 = 0.3411 Q_R$$

$$a_6 = 0.41606094$$

$$\alpha_6 = 0.2558 Q_R$$

$$N_{\Delta Y_O} = 0.10688 Q_R^2$$

$$\alpha_1^*(0) = 0.0306 Q_R$$

$$\alpha_2^*(0) = 0.0212 Q_R$$

$$\alpha_3^*(0) = 0.0821 Q_R$$

$$\alpha_4^*(0) = -0.0164 Q_R$$

$$\alpha_5^*(0) = -0.0339 Q_R$$

$$\alpha_6^*(0) = -0.1192 Q_R$$

$$N_{\epsilon_O} = 0.09896 Q_R^2$$

Beispiel 6:

$$r = 6 \quad (4) \quad a_1 = 0.52855938$$

$$\alpha_1 = -0.3444 Q_R$$

$$a_2 = -0.20002969$$

$$\alpha_2 = 0.3962 Q_R$$

$$a_3 = 0.0$$

$$\alpha_3 = 0.0$$

$$a_4 = 0.0$$

$$\alpha_4 = 0.0$$

$$a_5 = -0.12233516$$

$$\alpha_5 = 0.3411 Q_R$$

$$a_6 = 0.41606094$$

$$\alpha_6 = 0.2558 Q_R$$

$$\alpha_1^*(0) = 0.0306 Q_R$$

$$N_{\Delta Y_O} = 0.10454 Q_R^2$$

$$\alpha_2^*(0) = 0.0212 Q_R$$

$$\alpha_3^*(0) = 0.0$$

$$\alpha_4^*(0) = 0.0$$

$$N_{\epsilon_O} = 0.09896 Q_R^2$$

$$\alpha_5^*(0) = -0.0339 Q_R$$

$$\alpha_6^*(0) = -0.1192 Q_R$$

5. ENSEMBLE-UNTERSUCHUNGEN

5.1 Zweck der Ensemble-Betrachtungen

Mit den in Kapitel 4 abgeleiteten Gleichungen ist jedes individuelle System, das mit weißen, gleichverteilten Rauschsignalen erregt wird, in seinem Fehlerverhalten zu analysieren. Bei der Untersuchung einzelner Beispiele in Abschnitt 4.7 wurden bereits einige typische Fehlereigenschaften erkennbar; z.B. die Reduktion der Rest-Koeffizientenfehler oder die Annäherung der Fehlerleistung an die Werte nach Peled und Liu für eine große Anzahl von Koeffizienten.

In diesem Kapitel soll nun versucht werden, zu quantitativen Aussagen über die im Mittel zu erwartenden Fehlergrößen zu kommen. Dabei interessieren beispielsweise Erwartungswerte und Varianzen der Fehlerleistung, Verteilungen der Rest-Koeffizientenfehler, deren Varianzen usw. Es sind dazu statistische Betrachtungen über verschiedene Koeffizientensätze durchzuführen, die von der Statistik im Sinne der vorangegangenen Kapitel, also der Betrachtung des zeitlichen Verlaufs einzelner Signale zu unterscheiden sind. Zur Kennzeichnung der Koeffizienten-Statistik werden deshalb andere Symbole verwendet:

Erwartungswert:

$$E \{N_{\Delta Y}\} = \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{\ell=1}^L N_{\Delta Y}^{(\ell)} \quad (5.1a)$$

Varianz:

$$\text{VAR} \{N_{\Delta Y}\} = \lim_{L \rightarrow \infty} \frac{1}{L-1} \sum_{\ell=1}^L (N_{\Delta Y}^{(\ell)} - E\{N_{\Delta Y}\})^2 \quad (5.1b)$$

Der hochgestellte Index ℓ bezeichnet dabei die Fehlerleistung des individuellen Koeffizientensatzes ℓ .

Zunächst werden in Abschnitt 5.2 für den Fall $r = 2$ analytische Ausdrücke abgeleitet. Bei dem Versuch einer analytischen Formulierung für höhere Koeffizientenanzahl stößt man auf mathema-

tische Schwierigkeiten. Es werden deshalb in den nachfolgenden Kapiteln große Ensembles von Koeffizientensätzen auf dem Digitalrechner durchgerechnet, woraus die gesuchten statistischen Aussagen gewonnen werden können.

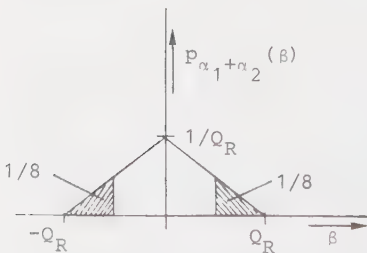
5.2 Analytische Formulierung für $r \equiv 2$

5.2.1 Zweierkomplement

In Abschnitt 4.7.2 wurde aufgezeigt, daß im Falle von zwei Koeffizienten prinzipiell zwei Möglichkeiten unterschieden werden müssen:

- a) Der lineare Fall bei $e_3^{(0)} = 0$, d.h. $|\alpha_1 + \alpha_2| < \frac{Q_R}{2}$, führt auf $\alpha_{1,2}^{*(0)} = \alpha_{1,2}$ und $N_{\epsilon_0} = 0$.
- b) Der nichtlineare Fall bei $e_3^{(0)} = \pm 1$, d.h. $|\alpha_1 + \alpha_2| \geq \frac{Q_R}{2}$, führt auf $\alpha_{1,2}^{*(0)} = \alpha_{1,2} \pm Q_R/2$ und $N_{\epsilon_0} = Q_R^2/12$.

Dabei tritt bei Annahme der Unabhängigkeit von α_1 und α_2 und einer jeweiligen Gleichverteilung im Intervall $\left[-\frac{Q_R}{2}, \frac{Q_R}{2}\right]$ der nichtlineare Fall mit einer Wahrscheinlichkeit $1/4$ auf, wie sich direkt aus der Faltung der beiden Verteilungsdichte-Funktionen p_{α_1} und p_{α_2} ergibt (siehe Bild 5.1).



$$\begin{aligned}
 P(|\alpha_1 + \alpha_2| \geq Q_R/2) &= \\
 &= \int_{-Q_R}^{-Q_R/2} p_{\alpha_1 + \alpha_2}(\beta) d\beta + \int_{Q_R/2}^{Q_R} p_{\alpha_1 + \alpha_2}(\beta) d\beta = \\
 &= 1/4
 \end{aligned}$$

Bild 5.1: Verteilung der Summe von zwei Koeffizientenfehlern

Der erste Schritt der analytischen Formulierung der Koeffizienten-Statistik ist die Berechnung der Verteilung von α_1^* bzw. α_2^* . Dazu wird zunächst die Verteilung derjenigen Koeffizienten α_{nl} ermittelt, die auf nichtlineare Fälle führen. Nach einiger Rechnung (siehe Anhang 6) erhält man den in Bild 5.2 dargestellten Verlauf.

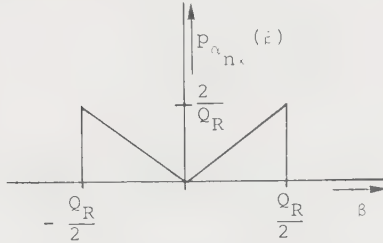


Bild 5.2: Verteilung der Koeffizientenfehler, die auf nichtlineare Fälle führen

Die Gleichverteilung $p_\alpha(\beta)$ der Koeffizientenfehler kann damit gemäß Bild 5.3a,b,c in zwei Teilfunktionen aufgespalten werden:

$$p_\alpha(\beta) = p_\alpha^{(1)}(\beta) + p_\alpha^{(2)}(\beta) \quad \text{mit} \quad (5.2a)$$

$$p_\alpha^{(1)}(\beta) = \frac{1}{4} \cdot p_{\alpha_{nl}}(\beta)$$

wegen der Wahrscheinlichkeit $\frac{1}{4}$ der nichtlinearen Fälle.

Daraus erhält man die Verteilung von $\alpha_{1,2}^*$, wenn man berücksichtigt, daß bei linearen Fällen ($p_\alpha^{(2)}(\beta)$) die a-priori-Koeffizientenfehler erhalten bleiben, während diese bei nichtlinearen Fällen ($p_\alpha^{(1)}(\beta)$) um $Q_R/2$ verschoben werden (Bild 5.3d). Die Summe der Funktionen nach Bild 5.3c und d ergibt $p_{\alpha_1^*}(\beta) = p_{\alpha_2^*}(\beta)$ Bild (5.3e).

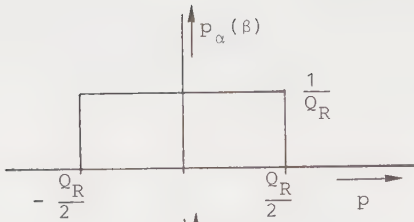


Bild 5.3a

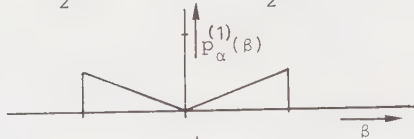


Bild 5.3b

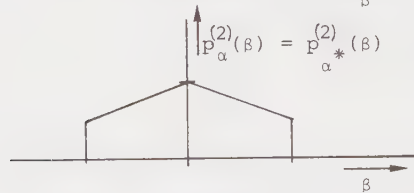


Bild 5.3c

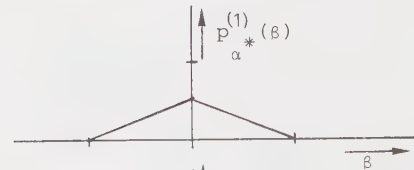


Bild 5.3d

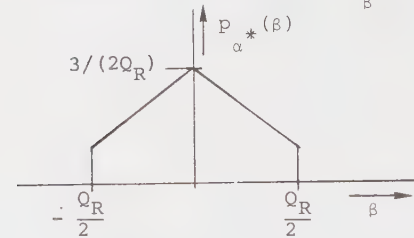


Bild 5.3e

Bild 5.3a-e: Konstruktion der Verteilung der Rest-Koeffizientenfehler

Damit erhält man die gesuchten statistischen Aussagen für $r = 2$:

Varianz der Rest-Koeffizientenfehler:

$$\text{VAR}(\alpha_{1,2}^*) = \int_{-Q_R/2}^{Q_R/2} p_{\alpha^*}(\beta) \cdot \beta^2 d\beta = \frac{Q_R^2}{16} \quad (5.3)$$

Erwartungswert der unkorrelierten Fehlerleistung:

$$E\{N_{\varepsilon_O}\} = \frac{1}{4} \cdot \frac{Q_R^2}{12} = \frac{Q_R^2}{48} \quad (5.4)$$

Erwartungswert der Gesamtleistung:

$$\begin{aligned} E\{N_{\Delta Y_O}\} &= 2 \cdot E\{(\alpha_{1,2}^{*(O)})^2 \sigma_x^2\} + E\{N_{\varepsilon_O}\} = \\ &= \frac{2}{3} \cdot \frac{Q_R^2}{16} + \frac{Q_R^2}{48} = \frac{Q_R^2}{16} \end{aligned} \quad (5.5)$$

5.2.2 Modifiziertes Zweierkomplement

Entsprechend den Überlegungen im letzten Abschnitt wird die Verteilung derjenigen Koeffizientenfehler ermittelt, für die gilt

$$|\alpha_1 + \alpha_2| \geq \frac{Q_R}{2} \quad \text{oder} \quad |\alpha_1 - \alpha_1| \geq \frac{Q_R}{2},$$

was in der Hälfte aller Fälle erfüllt ist.

Daraus wird wieder eine Aufteilung der Verteilungsdichtefunktion $p_{\alpha}(\beta)$ vorgenommen:

$$\begin{aligned} p_{\alpha}^{(1)}(\beta) \quad &\text{mit} \quad \alpha_{1,2}^{*(1)} = \alpha_{1,2} \pm \frac{Q_R}{2} \quad \text{und} \\ p_{\alpha}^{(2)}(\beta) \quad &\text{mit} \quad \alpha_{1,2}^{*(1)} = \alpha_{1,2} \quad (\text{vgl. Abschnitt 4.7.2}) \end{aligned}$$

Gemäß Bild 5.4a...e ergibt sich schließlich die Verteilung der Rest-Koeffizientenfehler.

Für das modifizierte Zweierkomplement erhält man damit:

$$\text{VAR}\{\alpha_{1,2}^{*(1)}\} = \frac{1}{24} \cdot Q_R^2 \quad (5.6)$$

$$E\{N_{\varepsilon_O}\} = 0 \quad (5.7)$$

$$E\{N_{\Delta Y_1}\} = \frac{1}{36} Q_R^2 \quad (5.8)$$

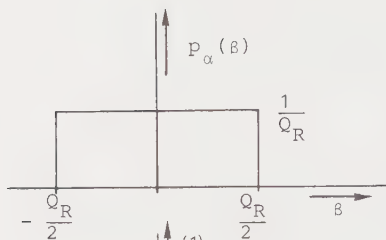


Bild 5.4a

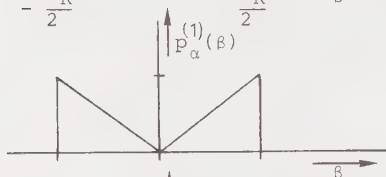


Bild 5.4b

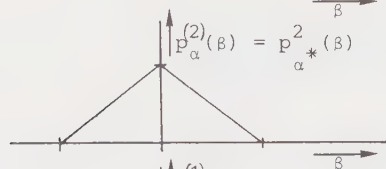


Bild 5.4c

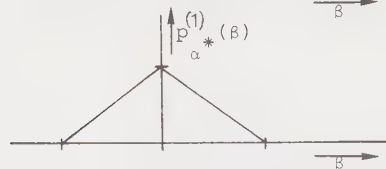


Bild 5.4d

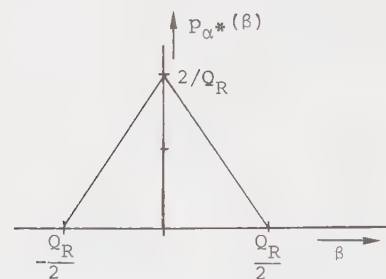


Bild 5.4e

Bild 5.4a...e: Konstruktion der Rest-Koeffizientenfehler-
Verteilung bei der modifizierten Form ($r=2$)

5.3 Untersuchung der Fehlerleistung

5.3.1 Erwartungswerte

Mit Hilfe eines hinreichend großen Ensembles von Werten $N_{\Delta Y}$, N_{ϵ} lassen sich deren Erwartungswerte näherungsweise nach den Gleichungen (5.9a und b) berechnen.

$$E\{N_{\Delta Y}\} = \frac{1}{L} \sum_{\ell=1}^L N_{\Delta Y}^{(\ell)} \quad \text{bzw.} \quad (5.9a)$$

$$E\{N_{\epsilon}\} = \frac{1}{L} \sum_{\ell=1}^L N_{\epsilon}^{(\ell)} \quad (5.9b)$$

Zur Ausführung der Mittelung würde ein Rechnerprogramm geschrieben. Dabei liefern r unabhängige Zufallsgeneratoren insgesamt L Koeffizientenfehler-Sätze $\mathbf{a}^1 \dots \mathbf{a}^L$; die r Koeffizientenfehler sind jeweils gleichverteilt im Intervall $\left[-\frac{Q_R}{2}, \frac{Q_R}{2}\right]$. Für jede dieser Stichproben werden nach den Gleichungen (4.54, 4.59, 4.68a und b) die Leistungen $N_{\Delta Y}^{(\ell)}$ und $N_{\epsilon}^{(\ell)}$ berechnet. Als Stichprobenmenge werden in den folgenden Untersuchungen jeweils $L = 50\,000$ Koeffizientensätze benutzt. Bild 5.5 zeigt die Ergebnisse für gewöhnliches und modifiziertes Zweierkomplement als Funktion von r .

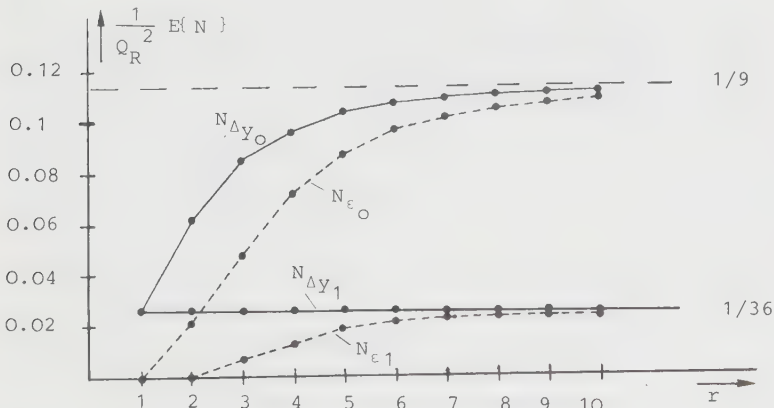


Bild 5.5: Erwartungswerte der Fehlerleistungen

Der Fall $r = 1$ entspricht der Konzentrierten Arithmetik; die Leistungen der unkorrelierten Fehleranteile verschwinden deshalb, die Erwartungswerte der Gesamtfehlerleistung bei gewöhnlichem und modifiziertem Zweierkomplement ergeben die gleichen Ausdrücke:

$$E\{N_{\Delta y}\} \Big|_{r=1} = \frac{1}{3} E\{\alpha^2\} = \frac{Q_R^2}{36} . \quad (5.10)$$

Für die modifizierte Form wurde in Abschnitt 5.2.2 analytisch derselbe Wert für $r = 2$ gefunden; außerdem entspricht er dem Grenzwert nach dem einfachen Modell von Peled/Liu und wird sich deshalb auch bei großer Koeffizientenzahl ergeben. Bild 5.5 zeigt demgemäß einen für alle r konstanten Verlauf von $E\{N_{\Delta y_1}\} = \frac{Q_R^2}{36}$. In Anhang 7 sind einige analytische Betrachtungen wiedergegeben, die dieses nicht selbstverständliche Verhalten der modifizierten Form näher erläutern sollen.

Der Erwartungswert der Gesamtfehlerleistung bei Zweierkomplement-Codierung steigt beginnend mit $Q_R^2/36$ bis zum Grenzwert $Q_R^2/9$ an, der bei $r = 9$ mit 98 % annähernd erreicht ist.

Die Leistungen der unkorrelierten Fehleranteile nähern sich mit steigendem r den Kurven der Gesamtleistungen: Für $r = 10$ bestehen beim gewöhnlichen Zweierkomplement 98,8 %, beim modifizierten 96,0 % der Gesamtfehlerleistung aus unkorreliertem Rauschen. Bild 5.6 zeigt das Leistungsverhältnis von unkorreliertem Fehler zu Gesamtfehler als Funktion von r . Der prozentuale Leistungsanteil der unkorrelierten Fehlergröße ist danach bei der modifizierten Form für alle r geringer als im gewöhnlichen Zweierkomplement.

Schließlich sind in Bild 5.7 die Erwartungswerte der auf r bezogenen Fehlerleistungen aufgetragen. Diese Größen sind deshalb von Interesse, weil aus Speicherplatzgründen oftmals das innere Produkt von Vektoren der Länge R aus R/r Teilprodukten mit jeweils einer Komponentenanzahl r zusammengesetzt werden muß (siehe Abschnitt 3.2). In diesem Falle ist der Erwartungswert der Fehlerleistung

$$E\{N_{\Delta Y}\} \Big|_R = R \cdot \left(\frac{1}{r} E\{N_{\Delta Y}\} \Big|_r \right) \quad (5.11a)$$

bzw. $E\{N_{\epsilon}\} \Big|_R = R \cdot \left(\frac{1}{r} E\{N_{\epsilon}\} \Big|_r \right) \quad (5.11b)$

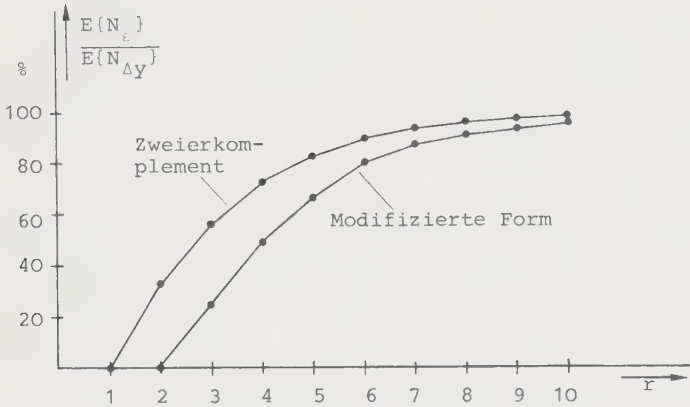


Bild 5.6: Verhältnis von unkorrelierter Rauschleistung zu Gesamtfehlerleistung

Es stellt sich dabei heraus, daß beim Zweierkomplement mit $r = 2$ der ungünstigste Fall vorliegt: Für größere Vektorlängen der Teilprodukte, aber auch für die Konzentrierte Arithmetik ($r = 1$) ergeben sich im Mittel geringere Gesamtfehlerleistungen am Ausgang. Im Gegensatz dazu fällt bei der modifizierten Form der Erwartungswert der resultierenden Fehlerleistung stets mit steigendem r .

Die maximalen Erwartungswerte der unkorrelierten Fehleranteile ergeben sich beim Zweierkomplement für $r = 4$, bei der modifizierten Form für $r = 5$.

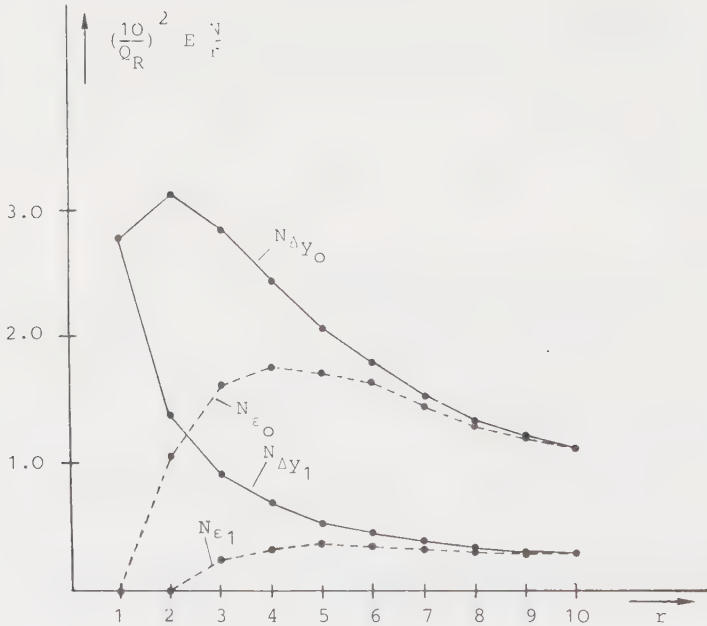


Bild 5.7: Erwartungswerte der bezogenen Fehlerleistungen

5.3.2 Varianzen

Bei der Untersuchung von Einzel-Beispielen in Abschnitt 4.7 wurde bereits deutlich, daß bei kleiner Anzahl von Koeffizienten die Fehlerleistungs-Werte in Abhängigkeit der individuellen Koeffizientensätze stark streuen, während für große r die Erwartungswerte ziemlich genau getroffen werden. Um darüber quantitative Aussagen zu finden, werden die Varianzen der Fehlerleistungen aus Ensembles von $L = 50\,000$ Koeffizientenfehler-Sätzen berechnet.

$$\text{VAR} \{N_{\Delta Y}\} \approx \frac{1}{L-1} \sum_{\ell=1}^L (N_{\Delta Y}^{(\ell)} - E\{N_{\Delta Y}\})^2 \quad (5.12a)$$

bzw.

$$\text{VAR} \{N_{\epsilon}\} \approx \frac{1}{L-1} \sum_{\ell=1}^L (N_{\epsilon}^{(\ell)} - E\{N_{\epsilon}\})^2 \quad (5.12b)$$

Bild 5.8a zeigt die Ergebnisse als Funktion von r .

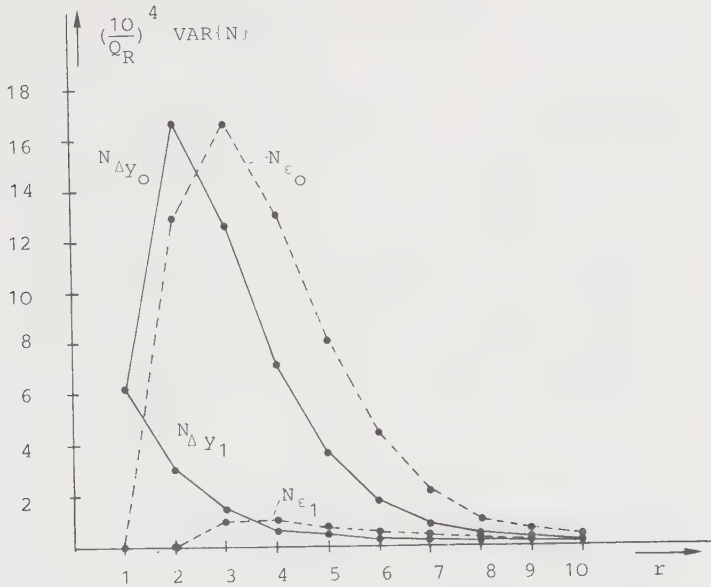


Bild 5.8a: Varianzen der Fehlerleistungen

Bei genauer Betrachtung erkennt man eine einfache Gesetzmäßigkeit für die modifizierte Form: Bei Erhöhung von r um 1 reduziert sich die Varianz der Gesamtfehlerleistung um den Faktor 2, was bei logarithmischer Skalierung der Ordinate deutlich wird (Bild 5.8b). Im gleichen Diagramm zeigt sich für die Gesamtfehler-Varianz der Zweierkomplement-Codierung das gleiche Verhalten für Werte $r \geq 4$; es gilt dabei

$$\text{VAR}\{N_{\Delta Y_0}\} / \text{VAR}\{N_{\Delta Y_1}\} \approx 9 \quad \text{für} \quad r \geq 4$$

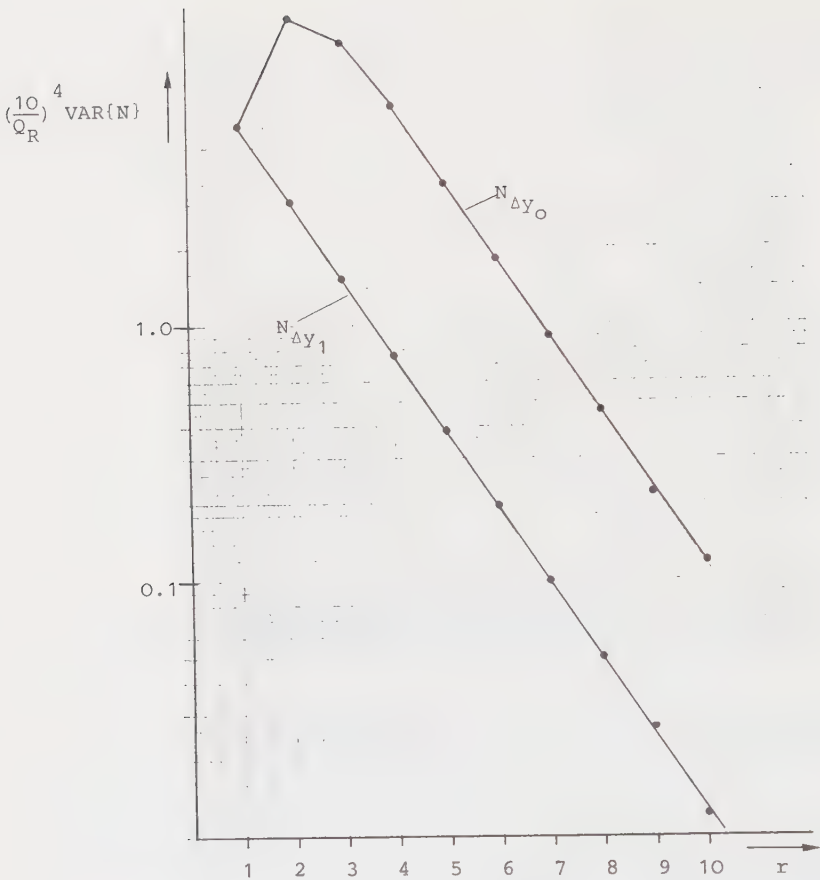


Bild 5.8b: Varianzen der Fehlerleistung bei logarithmischer Ordinaten-Skalierung

5.3.3 Verteilungen

Die Verteilungen der Fehlerleistungen werden nach folgenden Gleichungen bestimmt:

$$P_{N_{\Delta Y}}(\beta) = \frac{P(\beta \leq N_{\Delta Y} < \beta + \Delta\beta)}{\Delta\beta} \quad (5.13a)$$

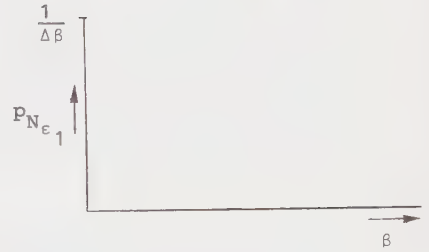
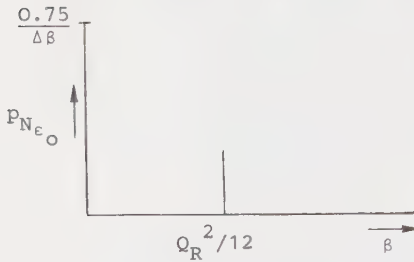
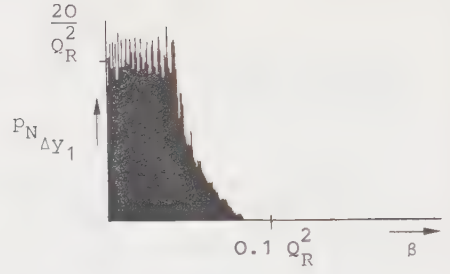
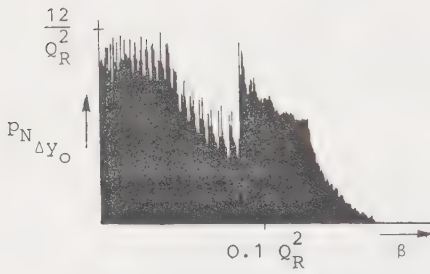
$$P_{N_{\epsilon}}(\beta) = \frac{P(\beta \leq N_{\epsilon} < \beta + \Delta\beta)}{\Delta\beta} \quad (5.13b)$$

mit
$$\Delta\beta = \frac{0.1 Q_R^2}{256} \quad (5.13c)$$

Dazu wird jeweils ein Ensemble von 50 000 Koeffizientenfehlersätzen zugrundegelegt.

In den Bildern 5.9a ... 5.9f sind die Ergebnisse wiedergegeben, und zwar jeweils für gewöhnliches und modifiziertes Zweierkomplement im Vergleich. Bei den Diagrammen der Gesamtfehler-Verteilungen (jeweils die oberen Bilder) handelt es sich um kontinuierliche Verteilungen, denn in den Berechnungsformeln für N_{AY} (Gl. 4.54, 4.59) tauchen die Werte der a-priori-Koeffizientenfehler auf, also kontinuierlich verteilte Größen. Für große Werte von r konzentrieren sich die Fehlerleistungen um die Werte $Q^2/9$ bzw. $Q^2/36$, wie gemäß des Peled/Liu-Modells zu erwarten ist.

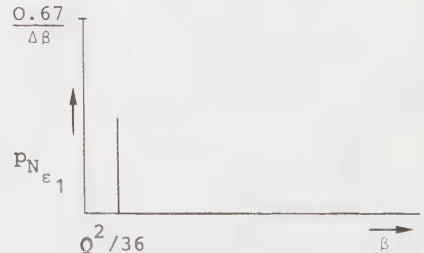
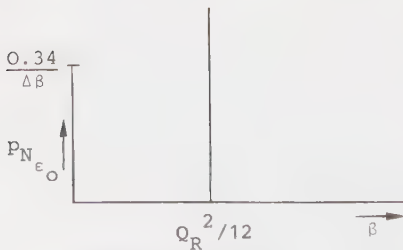
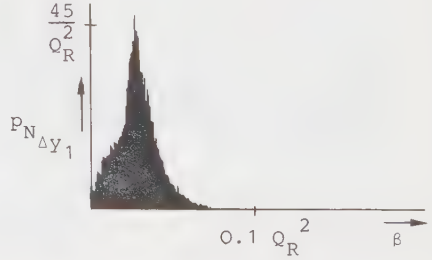
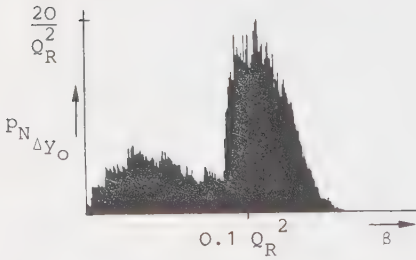
Die Verteilungen der unkorrelierten Fehlerleistung müssen diskret sein, da in den Gleichungen (4.68a,b) nur noch Ausdrücke auftreten, die die Komponenten der Fehlervektoren enthalten, und diese können nach Gl. (4.29) nur endlich viele Werte annehmen. Für wachsendes r wird die Anzahl der möglichen Werte N_e immer größer, so daß die Linien der diskreten Verteilungen immer enger zusammenrücken. Gleichzeitig konzentrieren sie sich um die Werte $Q^2/9$ bzw. $Q^2/36$ und gleichen sich damit schließlich den Verteilungen der Gesamtfehlerleistungen an.



Zweierkomplement

modifizierte Form

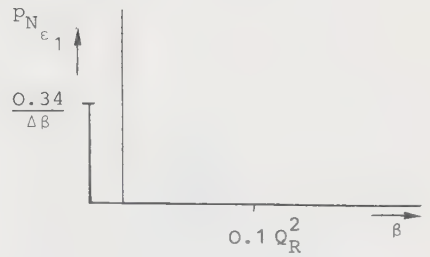
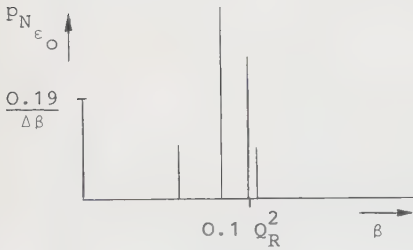
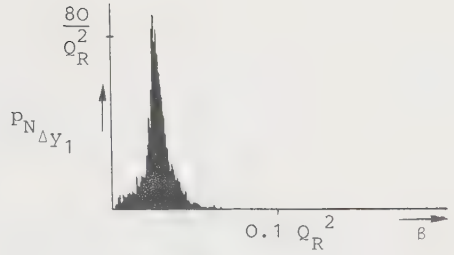
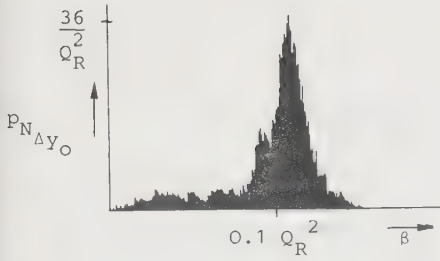
Bild 5.9a: Verteilungen der Fehlerleistungen bei $r=2$



Zweierkomplement

modifizierte Form

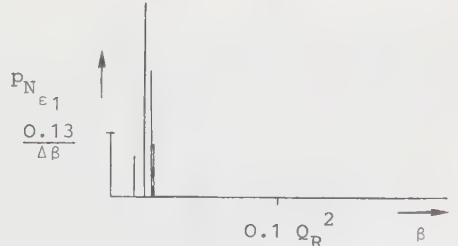
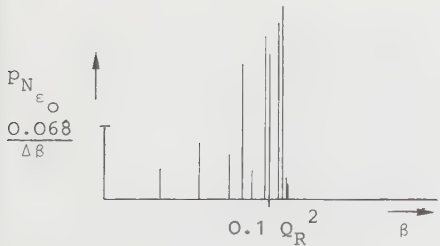
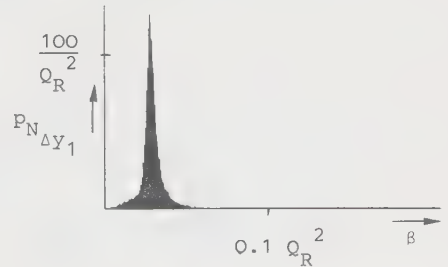
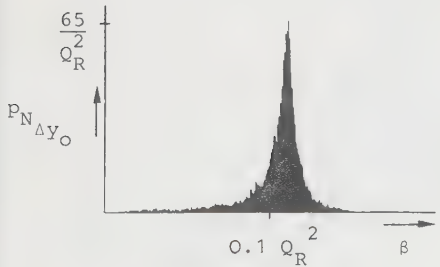
Bild 5.9b: Verteilungen der Fehlerleistungen bei $r=3$



Zweierkomplement

Modifizierte Form

Bild 5.9c: Verteilungen der Fehlerleistungen bei $r=4$



Zweierkomplement

Modifizierte Form

Bild 5.9d: Verteilungen der Fehlerleistungen bei $r=5$

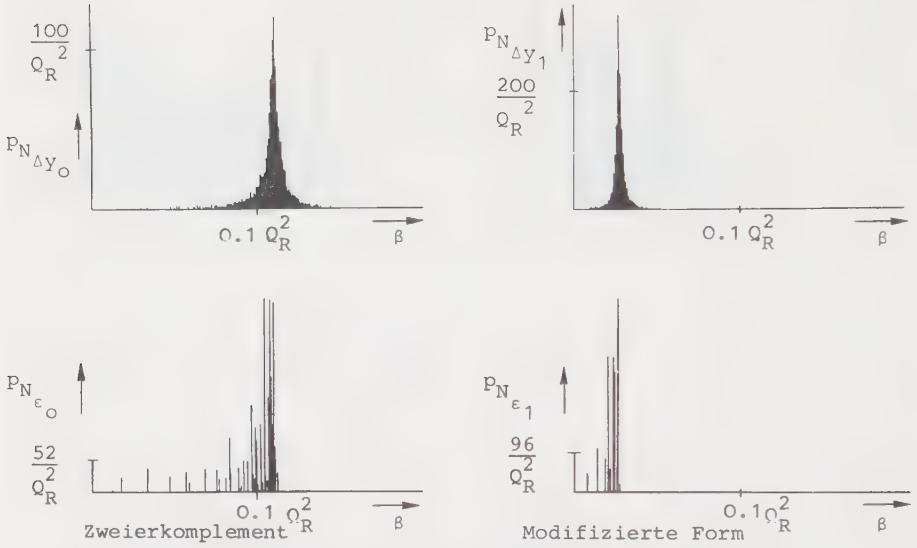


Bild 5.9e: Verteilungen der Fehlerleistungen bei $r=6$

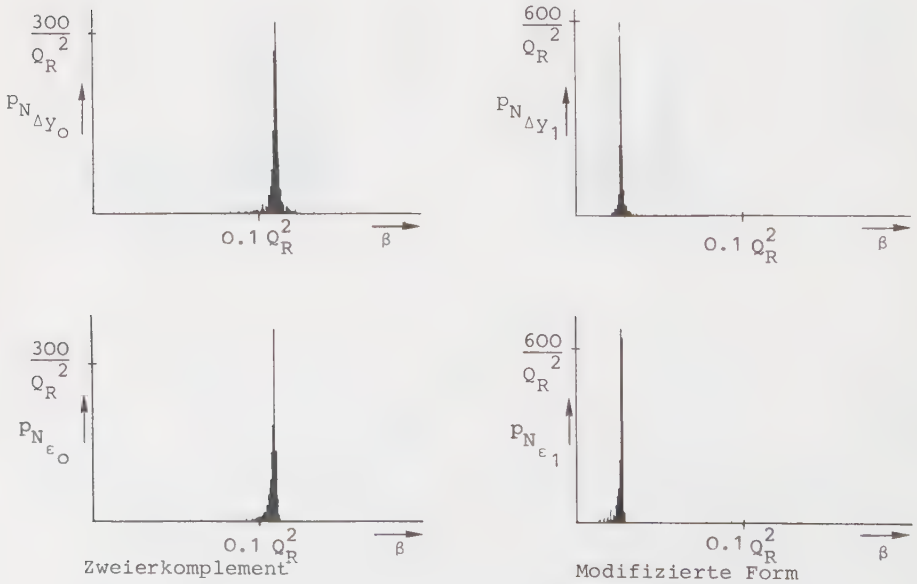


Bild 5.9f: Verteilungen der Fehlerleistungen bei $r=8$

5.4 Untersuchung der Rest-Koeffizientenfehler

5.4.1 Varianzen

Die Erwartungswerte der Rest-Koeffizientenfehler müssen aus Symmetriegründen verschwinden. Ihre Varianzen berechnen sich damit nach Gl. (5.14) aus einem Ensemble:

$$\text{VAR } \{\alpha_{\rho}^*\} = E\{(\alpha_{\rho}^*)^2\} \approx \frac{1}{L-1} \sum_{\ell=1}^L (\alpha_{\rho}^{*\ell})^2 \quad (5.14)$$

Da für die r Eingangssignale x_{ρ} einerseits und die r a-priori-Koeffizientenfehler andererseits jeweils gleiche Verteilungen und statistische Unabhängigkeit untereinander angenommen wurde, liegen für alle r Rest-Koeffizientenfehler statistisch die gleichen Verhältnisse vor. Es genügt deshalb, Gl. (5.14) für einen beliebigen Restkoeffizientenfehler α_{ρ}^* , $\rho \in \{1, 2, \dots, r\}$, auszuwerten, der stellvertretend für alle übrigen steht. Die Berechnung der Werte $\alpha_{\rho}^{*\ell}$ für die einzelnen Stichproben ℓ erfolgt nach Gl. (4.66).

Die Darstellung der Ergebnisse in Bild 5.15a zeigt die erwarteten Verläufe: Beginnend mit der Varianz der gleichverteilten a-priori-Koeffizientenfehler $Q_R^2/12$ bei der Konzentrierten Arithmetik ($r=1$) fallen beide Kurven mit wachsendem r ab; dabei die der modifizierten Form schneller als die des gewöhnlichen Zweierkomplements.

Bei logarithmischer Skalierung der Ordinate (Bild 5.15b) findet man für die modifizierte Form eine Eigenschaft, die an die Betrachtungen der Fehlerleistungs-Varianzen erinnert (vgl. Abschnitt 5.3.2): Bei Erhöhung des Wertes r um 1 halbiert sich jeweils die Varianz der Rest-Koeffizientenfehler. Die entsprechende Kurve für das gewöhnliche Zweierkomplement verläuft ab $r=4$ etwa parallel und hat dabei etwa den doppelten Wert.

Für die Varianzen der Restkoeffizientenfehler können damit empirisch gefundene Formeln angegeben werden:

$$\text{VAR} \{ \alpha_{\rho}^{*(0)} \} \approx \frac{Q_R^2}{12} 2^{-r+2} \text{ für } r \geq 4 \quad (5.15a)$$

$$\text{VAR} \{ \alpha_{\rho}^{*(1)} \} = \frac{Q_R^2}{12} \cdot 2^{-r+1} \quad \forall r \quad (5.15b)$$

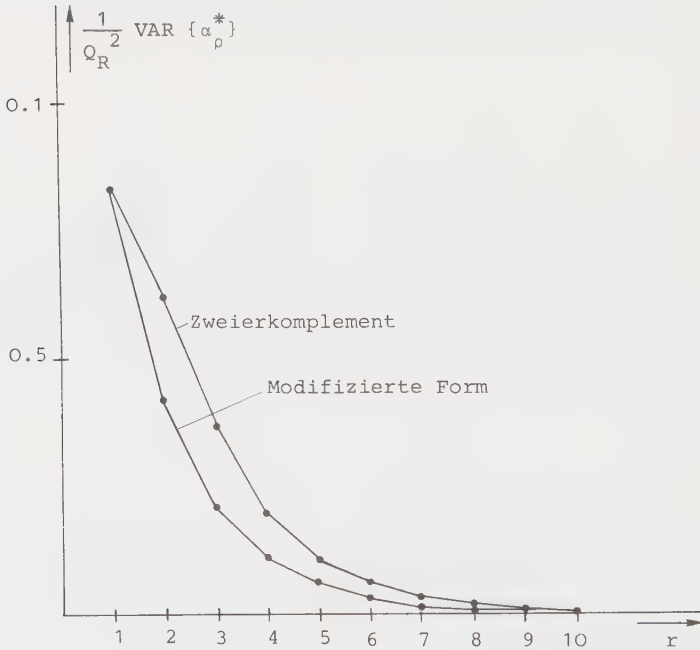


Bild 5.10a: Varianzen der Rest-Koeffizientenfehler

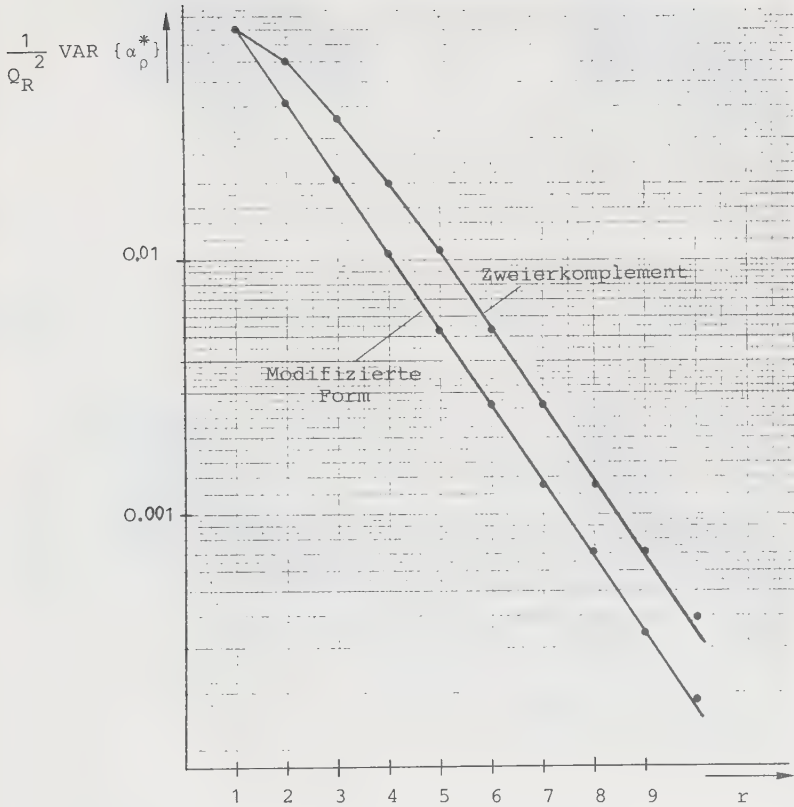


Bild 5.10b: Varianzen der Rest-Koeffizientenfehler
bei logarithmischer Skalierung

5.4.2 Verteilungen

Abschließend werden für verschiedene r die Verteilungen der Rest-Koeffizientenfehler aus Ensembles von je 50 000 Koeffizientenfehler-Sätzen bestimmt:

$$P_{|\alpha_p^*|}(\beta) = \frac{P(\beta \leq |\alpha_p^*| \leq \beta + \Delta\beta)}{\Delta\beta} ; \quad \Delta\beta = Q_R/512 . \quad (5.16a)$$

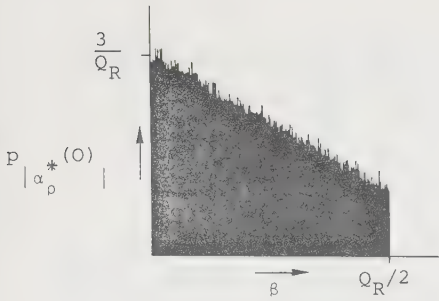
Wegen der zu erwartenden Symmetrie

$$P_{\alpha_p^*}(\beta) = P_{\alpha_p^*}(-\beta) \quad (5.16b)$$

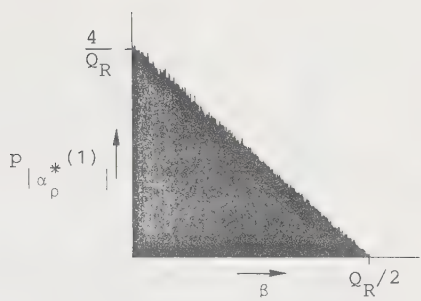
wird die Berechnung nach (5.16a) auf die Beträge $|\alpha_p^*|$ beschränkt, um so die Anzahl der Stichproben im Intervall $[0, \frac{Q_R}{2}]$ zu erhöhen.

In den Bildern 5.11a...f sind jeweils links die Verteilungen bei gewöhnlichem, rechts bei modifiziertem Zweierkomplement dargestellt. Für $r=2$ ergeben sich die analytisch bestimmten Verläufe (nach Abschnitt 5.2); bei Erhöhung von r werden die Verteilungen zunehmend schmaler, und zwar bei modifiziertem in stärkerem Maße als beim einfachen Zweierkomplement. Dabei sind alle Werte auf den Bereich $[0, \frac{Q_R}{2}]$ beschränkt, d.h. für die Rest-Koeffizientenfehler gilt bei Erregung mit r weißen, gleichverteilten und unabhängigen Rauschquellen für jedes r

$$-\frac{Q_R}{2} \leq \alpha_p^* \leq \frac{Q_R}{2} \quad (5.17)$$



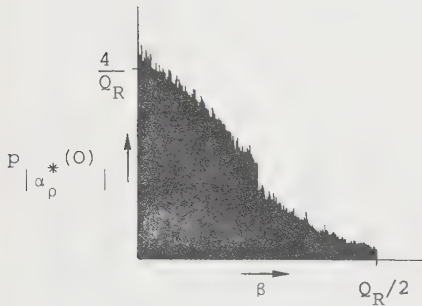
Zweierkomplement



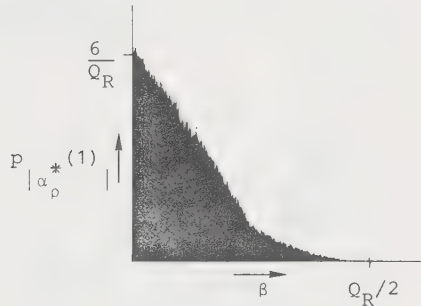
r=2

Modifizierte Form

Bild 5.11a: Verteilungen der Rest-Koeffizientenfehler



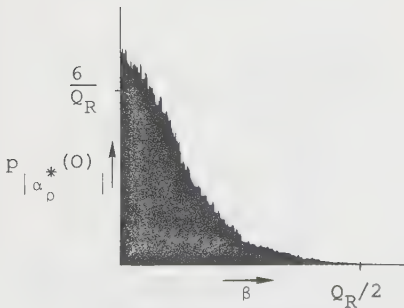
Zweierkomplement



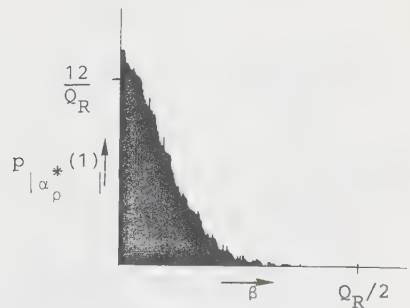
r=3

Modifizierte Form

Bild 5.11b: Verteilungen der Rest-Koeffizientenfehler



Zweierkomplement



r=4

Modifizierte Form

Bild 5.11c: Verteilungen der Rest-Koeffizientenfehler

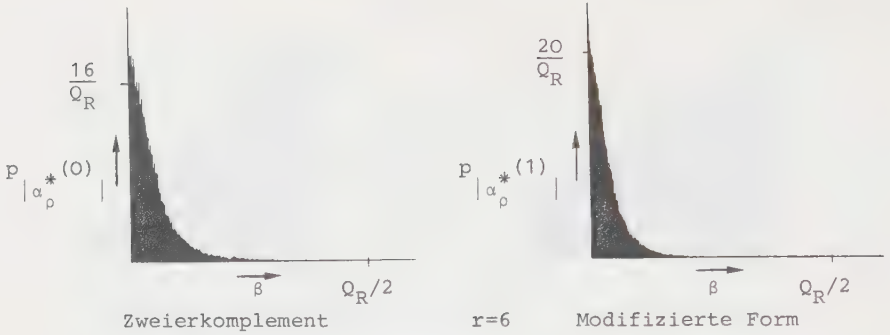


Bild 5.11d: Verteilungen der Rest-Koeffizientenfehler

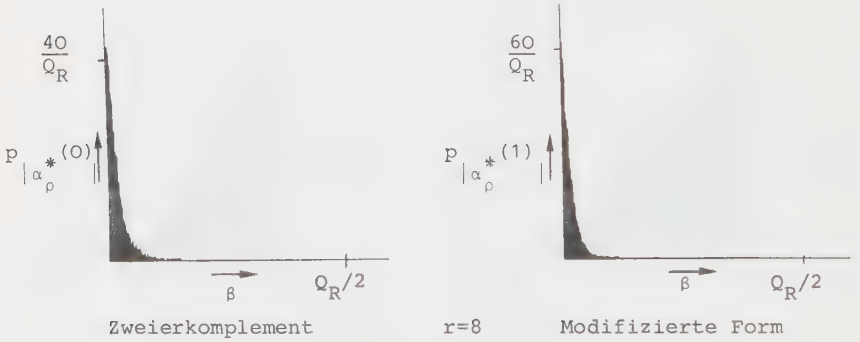
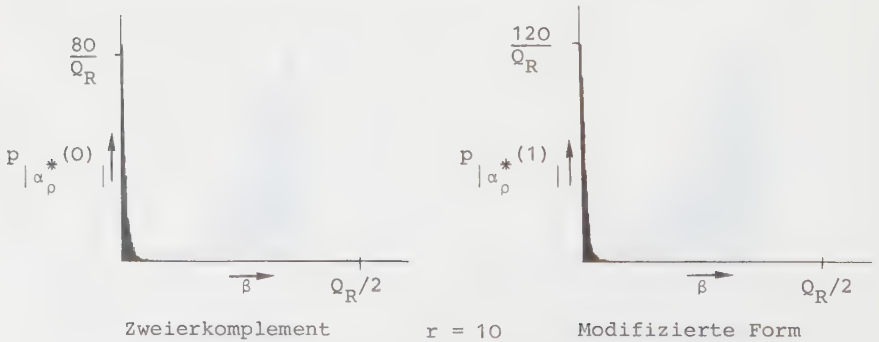


Bild 5.11e: Verteilungen der Rest-Koeffizientenfehler



5.11f: Verteilungen der Rest-Koeffizientenfehler

6. FEHLERANALYSE BEI UNKORRELIERTEN VARIABLEN MIT BELIEBIGEN STATISTISCHEN EIGENSCHAFTEN

6.1 Simulation einiger Beispiele

Es soll zunächst durch Simulation geklärt werden, inwieweit die in Kapitel 4 abgeleitete Fehleranalyse für weiße, gleichverteilte Eingangssignale auch bei Erregung mit farbigem Rauschen ihre Gültigkeit behält.

Dazu wird wieder das Simulations-Blockdiagramm nach Bild 4.8 verwendet, mit dem Unterschied, daß hier die Ausgänge der r weißen Rauschquellen zur Spektralformung jeweils über ein Filter mit der Übertragungsfunktion

$$H_f(z) = \frac{b z^2}{z^2 + c_1 z + c_0} \quad (6.1)$$

geführt werden. Da hier zunächst der Spezialfall statistisch gleichartiger Eingangssignale betrachtet wird, werden die Koeffizienten der r Spektralformungs-Filter gleich gewählt. Die Leistungsspektren der Signale x_ρ sind damit identisch:

$$|X_\rho(\Omega)|^2 = \frac{1}{3} \cdot |H_f(\Omega)|^2, \quad \rho \in \{1, \dots, r\} \quad (6.2)$$

Dabei bleiben die $x_\rho(k)$ weiterhin unkorreliert wegen der Ansteuerung der r Filter durch unkorrelierte Rauschquellen.

Als Beispiel wird hier für die Spektralformungs-Filter willkürlich angenommen:

$$\text{Polradius} \quad \rho_\infty = 0.9 \rightarrow c_0 = 0.81 \quad (6.3a)$$

$$\text{Polwinkel} \quad \varphi_\infty = 0.3\pi \rightarrow c_1 = 1.058013 \quad (6.3b)$$

Die Zählerkonstante b wird so bestimmt, daß nach 100 000 Abtastwerten an den Ausgängen der Filter gerade keine Übersteuerungen stattgefunden haben. Auf diese Weise liegt das entstehende Signal im Intervall $[-1, 1)$, entspricht also im Aussteuerungsbereich den im letzten Abschnitt benutzten gleich-

verteilten Rauschvariablen. Es ergibt sich der Wert

$$b = 0.21442 \quad (6.3c)$$

Bild 6.1 zeigt den Betragsfrequenzgang $|H_f(\Omega)|$ für dieses Beispiel.

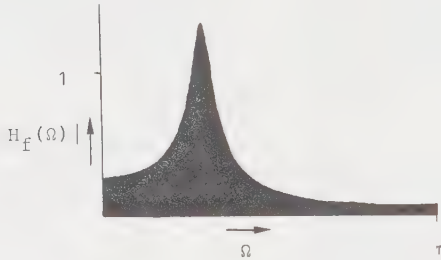


Bild 6.1: Frequenzgang der Spektralformungs-Filter

Als Koeffizientensätze $a_1, \dots, a_r$ werden die bereits aus Kap. 4 bekannten Beispiele herangezogen.

Simulationsergebnisse aus je 50 000 Stichproben ($Q_R=2^{-7}$)

Beispiel_1: ($r=2$)

$$\begin{aligned} a_1 &= 0.2178328 & \alpha_1 &= -0.1174 Q_R \\ a_2 &= -0.3159031 & \alpha_2 &= -0.4356 Q_R \end{aligned}$$

	Zweierkomplement	Modifizierte Form
$N_{\Delta Y}$	$0.03910 Q_R^2$	$0.01028 Q_R^2$
N_ϵ	$0.02874 Q_R^2$	0.0
α_1^*	$0.3850 Q_R$	$0.3860 Q_R$
α_2^*	$0.0651 Q_R$	$0.0649 Q_R$

Beispiel_2 (r=2):

$$\begin{aligned} a_1 &= 0.4941047 & \alpha_1 &= 0.2454 Q_R \\ a_2 &= 0.1852891 & \alpha_2 &= -0.2830 Q_R \end{aligned}$$

	Zweierkomplement	Modifizierte Form
$N_{\Delta Y}$	$0.00960 Q_R^2$	$0.00770 Q_R^2$
N_ϵ	0.0	0.0
α_1^*	$0.2456 Q_R$	$-0.2550 Q_R$
α_2^*	$-0.2831 Q_R$	$0.2175 Q_R$

Beispiel_3 (r=8):

$$\begin{aligned} a_1 &= 0.251531 & \alpha_1 &= 0.196 Q_R \\ a_2 &= 0.218546 & \alpha_2 &= -0.026 Q_R \\ a_3 &= -0.405625 & \alpha_3 &= 0.080 Q_R \\ a_4 &= 0.849391 & \alpha_4 &= -0.278 Q_R \\ a_5 &= -0.260927 & \alpha_5 &= -0.399 Q_R \\ a_6 &= 0.282398 & \alpha_6 &= 0.147 Q_R \\ a_7 &= 0.081344 & \alpha_7 &= 0.412 Q_R \\ a_8 &= 0.155047 & \alpha_8 &= 0.154 Q_R \end{aligned}$$

	Zweierkomplement	Modifizierte Form
$N_{\Delta Y}$	$0.0519 Q_R^2$	$0.0148 Q_R^2$
N_ϵ	$0.0515 Q_R^2$	$0.0147 Q_R^2$
α_1^*	$0.01631 Q_R$	$-0.02275 Q_R$
α_2^*	$-0.00256 Q_R$	$0.00525 Q_R$
α_3^*	$-0.02156 Q_R$	$0.01375 Q_R$

	Zweierkomplement	Modifizierte Form
α_4^*	-0.02020 Q_R	-0.01238 Q_R
α_5^*	-0.04744 Q_R	-0.00237 Q_R
α_6^*	-0.00144 Q_R	-0.00925 Q_R
α_7^*	0.04438 Q_R	0.00575 Q_R
α_8^*	0.00556 Q_R	-0.00224 Q_R

Im Vergleich zu den Analyse-Ergebnissen von Kapitel 4.7.1 zeigt sich überraschenderweise, daß die Restkoeffizientenfehler in allen Beispielen bei weißem und farbigem Eingangsauschauschen übereinstimmen. Demgegenüber stellt man große Abweichungen in den Leistungswerten fest, was nicht verwundern darf, denn durch die Filterung verändert sich die Leistung der Eingangssignale x_p gegenüber der der weißen Rauschquellen. Im vorliegenden Beispiel ist die Leistung des farbigen Signals:

$$\sigma_{x_f}^2 = 0.067968 .$$

Demgemäß wird die resultierende Fehlerleistung bei den linearen Systemen gegenüber den Ergebnissen von Kapitel 4.7.1 um den Faktor

$$\frac{1}{3} / \sigma_{x_f}^2 = 4.90426$$

reduziert.

	$N_{\Delta Y}$ nach Simulation	$\frac{1}{4.90426} \cdot N_{\Delta Y} _{\text{weiß}}$
Beispiel 1 (mod.Form)	0.01028 Q_R^2	0.0105622 Q_R^2
Beispiel 2 (mod.Form)	0.00770 Q_R^2	0.0076056 Q_R^2
Beispiel 2 (Zweierkomp.)	0.00960 Q_R^2	0.0095366 Q_R^2

In den anderen (nichtlinearen) Fällen ist ein derartiger einfacher Zusammenhang mit der Eingangsleistung nicht anzugeben: Die Untersuchung der Fehlerleistung erfordert also eine genauere Analyse.

6.2 Vorbetrachtungen zur Fehleranalyse

Betrachtet man die Voraussetzung für die Eingangssignale, auf denen die Analyse in Kap. 4 beruht, so findet man keine Bedingung bezüglich der Korrelation zeitlich aufeinanderfolgender Werte $x(k)$ und $x(k+\kappa)$. Mit anderen Worten: solange die einzelnen Rauschquellen unkorreliert sind, berühren die Autokorrelationsfunktionen der einzelnen Eingangssignale die Fehleranalyse nicht, und damit kann auch die spektrale Formung primär nicht für die Abweichung der Fehlerleistungswerte verantwortlich sein. Andererseits wurde jedoch in Kap. 4 eine gleichförmige Amplitudenverteilung für die x_p angenommen. Durch die Filterung der von den Rauschquellen gelieferten Signale ergibt sich i.a. eine wesentlich andere Verteilung, die sich der Normalverteilung nähert. Bild 6.2 macht dieses für das gewählte Beispiel der Spektralformung deutlich.

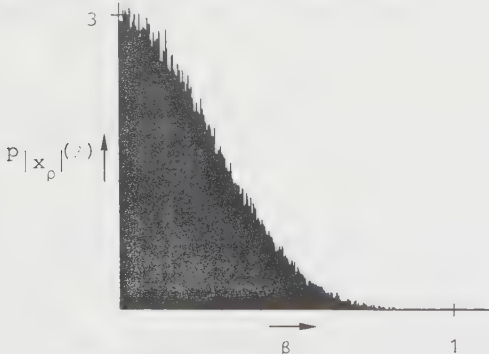


Bild 6.2: Verteilung von x_p :

$$p_{x_p}(\beta) = \frac{P\{\beta \leq x_p < \beta + \Delta\beta\}}{\Delta\beta} ; \Delta\beta = 1/256$$

$$\text{für } \beta \geq 0$$

Für die Beschreibung des Fehlerverhaltens der Verteilten Arithmetik sind die Korrelationen zwischen den einzelnen Bits der Variablen entscheidend (vgl. Voraussetzungen bei weißem, gleichverteiltem Rauschen, Gl. (4.43)).

Wegen der hier angenommenen Unabhängigkeit der Variablen gilt für die Bits *v e r s c h i e d e n e r* Variablen x_ρ und x_σ wie bisher

$$\langle \xi_{\rho i} \cdot \xi_{\sigma j} \rangle = \langle \xi_{\rho i} \rangle \cdot \langle \xi_{\sigma j} \rangle = \frac{1}{4} \quad \text{für } \rho \neq \sigma, \quad (6.4) \\ \forall i, j \in \{0, \dots, w-1\}$$

Die einzelnen Bits *e i n e r* Variablen hingegen sind nunmehr als korreliert zu betrachten, d.h. es ist

$$\langle \xi_{\rho i} \xi_{\rho j} \rangle \neq \langle \xi_{\rho i} \rangle \cdot \langle \xi_{\rho j} \rangle = \frac{1}{4} \quad \text{für } i \neq j. \quad (6.5)$$

Es sind also zur Formulierung des Fehlers die Werte $\langle \xi_{\rho i} \xi_{\rho j} \rangle$ zu bestimmen, die allein von der Verteilung des Signals $x_\rho(k)$ abhängen.

Man macht sich diesen Sachverhalt sehr einfach am Beispiel der beiden vorderen Bits $\xi_{\rho 0}$ und $\xi_{\rho 1}$ klar, die bei kleinen Werten $x_\rho(k)$ jeweils gleich sind (1 bei negativen, 0 bei positiven Werten). Da aber, wie Bild 6.2 zu entnehmen ist, kleine x_ρ -Werte häufiger auftreten, wird der Fall $\xi_{\rho 0} \neq \xi_{\rho 1}$ entsprechend selten vorliegen, so daß sich der Mittelwert $\langle \xi_{\rho 0} \cdot \xi_{\rho 1} \rangle$ zu einem größeren Wert als bei gleichverteilten Eingangssignalen verschiebt.

Festzuhalten bleibt hier also die wichtige Tatsache, daß das Fehlerverhalten bei Erregung mit *u n k o r r e l i e r t e n* Signalen beliebiger statistischer Eigenschaften nicht durch ihre spektrale Form sondern *a u s s c h l i e ß l i c h* durch die *A m p l i t u d e n v e r t e i l u n g* bestimmt wird. Die bei den Simulationsbeispielen im letzten Abschnitt angewendete Methode der Signalerzeugung über spektralformende Filter ist somit nur als eine von vielen Möglichkeiten zu betrachten, die gleichförmige Amplitudenverteilung der vorliegenden Rauschquellen zu verändern. In Kap. 7 wird

sich zeigen, daß die spektrale Verteilung erst dann auf das Fehlerverhalten einwirkt, wenn Korrelationen zwischen den Variablen infolge von Verzögerungsvorgängen vorliegen.

6.3 Einführung der Bit-Korrelations-Matrix

Zur analytischen Beschreibung des Fehlerverhaltens werden also die Korrelationen zwischen sämtlichen möglichen Bit-Paaren von einer Variablen benötigt.

Es wird dazu eine Matrix definiert, die als "Bit-Korrelations-Matrix" bezeichnet werden soll:

$$\Phi = \begin{bmatrix} \psi_{00} & \psi_{01} & \dots & \psi_{0,w-1} \\ \psi_{10} & \psi_{11} & & \vdots \\ \vdots & & & \vdots \\ \psi_{w-1,0} & \dots & & \psi_{w-1,w-1} \end{bmatrix} \quad (6.6)$$

Ihre Elemente beinhalten die Mittelwerte der Produkte jeweils zweier Bits:

$$\psi_{ij} = \langle \xi_{\rho i} \cdot \xi_{\rho j} \rangle \quad (6.7)$$

Es gilt $\psi_{ij} = \psi_{ji}$, d.h. die Matrix ist symmetrisch.

Da die r Eingangssignale hier als statistisch gleichartig angenommen werden, braucht die Matrix Φ nicht mit ρ indiziert zu werden.

Die Bit-Korrelations-Matrix wird zunächst für das vorliegende Beispiel durch Simulation bestimmt (100 000 Werte, $w=8$).

$$\Phi = \begin{bmatrix} 0.500 & 0.474 & 0.353 & 0.296 & 0.273 & 0.263 & 0.254 & 0.254 \\ 0.474 & 0.501 & 0.329 & 0.291 & 0.265 & 0.259 & 0.253 & 0.253 \\ 0.353 & 0.329 & 0.502 & 0.262 & 0.242 & 0.252 & 0.249 & 0.250 \\ 0.296 & 0.281 & 0.252 & 0.499 & 0.249 & 0.250 & 0.249 & 0.250 \\ 0.273 & 0.265 & 0.249 & 0.249 & 0.498 & 0.249 & 0.250 & 0.249 \\ 0.263 & 0.259 & 0.252 & 0.250 & 0.249 & 0.500 & 0.251 & 0.251 \\ 0.254 & 0.253 & 0.249 & 0.249 & 0.250 & 0.251 & 0.250 & 0.249 \\ 0.254 & 0.253 & 0.250 & 0.250 & 0.249 & 0.251 & 0.249 & 0.500 \end{bmatrix} \quad (6.8)$$

Sie läßt sich jedoch auch wie folgt aus der Amplitudenverteilung von x_p direkt berechnen $|31|$, was eine erhebliche Einsparung an Rechenzeit bedeutet.

Wird das Signal x_p auf die Eingangs-Quantisierungsstufe $Q_E = 2^{-w+1}$ gerundet, so erhält man für die Amplituden eine diskrete Verteilung mit insgesamt 2^w Linien.

Es sei

$$p_\lambda = P\{[x_p]_{Q_E} = Q_E \cdot \lambda; \lambda \in \{-2^{w-1}, \dots, +2^{w-1}-1\}\} \quad (6.9)$$

die Wahrscheinlichkeit des Auftretens eines der 2^w möglichen Werte von $[x_p]_{Q_E}$. Jeder Zahl λ lassen sich dabei die w Bit $\xi_{p0}, \dots, \xi_{pw-1}$ der Zweierkomplement-Darstellung von $[x_p]_{Q_E} = \lambda Q_E$ zuordnen.

Bei der Berechnung von $\varphi_{ij} = \langle \xi_{pi} \xi_{pj} \rangle$ liefern nur die Fälle $\xi_{pi} = \xi_{pj} = 1$ Beiträge, d.h. es sind die zugehörigen Werte p_λ aufzuaddieren. Zur Erläuterung betrachte man den einfachen Fall $w = 3$

λ	ξ_{p0}	ξ_{p1}	ξ_{p2}	Beiträge zu:
-4	1	0	0	φ_{00}
-3	1	0	1	$\varphi_{00}, \varphi_{22}, \varphi_{02}$
-2	1	1	0	$\varphi_{00}, \varphi_{11}, \varphi_{01}$
-1	1	1	1	$\varphi_{00}, \varphi_{11}, \varphi_{22}, \varphi_{01}, \varphi_{02}, \varphi_{12}$
0	0	0	0	-
1	0	0	1	φ_{22}
2	0	1	0	φ_{11}
3	0	1	1	$\varphi_{11}, \varphi_{22}, \varphi_{12}$

Tabelle 6.1: Berechnung der Werte φ_{ij}

Der Tabelle 6.1 entnimmt man die Berechnungsvorschriften für die Elemente φ_{ij} , z.B.

$$\varphi_{00} = p_{-4} + p_{-3} + p_{-2} + p_{-1}$$

$$\varphi_{01} = p_{-2} + p_{-1}$$

$$\varphi_{02} = p_{-3} + p_{-1}$$

usw.

Allgemein sind bei einer Wortlänge w für jeden Wert φ_{ij} jeweils 2^{w-2} (für $i = j$) bzw. 2^{w-1} (für $i \neq j$) Wahrscheinlichkeiten aufzuaddieren.

Der in Kap. 4 behandelte Spezialfall der Gleichverteilung führt dabei mit $p_\lambda = 2^{-w} \forall \lambda$ auf

$$\varphi_{ij} = \begin{cases} 2^{w-2} 2^{-w} = 1/4 & \text{für } i \neq j \\ 2^{w-1} 2^{-w} = 1/2 & \text{für } i = j \end{cases} \quad (\text{vgl. Gl. 4.43})$$

Das hier angegebene Verfahren wurde programmiert. Unter Benutzung der durch Simulation ermittelten Amplitudenverteilung nach Bild 6.2 erhält man die Bit-Korrelations-Matrix (6.8b),

die im Rahmen der Simulationsgenauigkeit mit 6.8a übereinstimmt. Setzt man weiter anstelle der Verteilung nach Bild 6.2 eine Normalverteilung für x_p (mit $\sigma_x^2 = 0.067968$), so erhält man die Matrix nach Gl. (6.8c). Es zeigen sich keine nennenswerten Unterschiede zu Gl. (6.8b).

$$\Phi = \begin{bmatrix} 0.494 & 0.463 & 0.351 & 0.294 & 0.271 & 0.259 & 0.254 & 0.249 \\ 0.463 & 0.436 & 0.328 & 0.280 & 0.265 & 0.256 & 0.253 & 0.249 \\ 0.351 & 0.328 & 0.500 & 0.253 & 0.250 & 0.251 & 0.251 & 0.249 \\ 0.294 & 0.280 & 0.253 & 0.500 & 0.251 & 0.250 & 0.251 & 0.249 \\ 0.271 & 0.265 & 0.250 & 0.251 & 0.500 & 0.250 & 0.252 & 0.249 \\ 0.259 & 0.256 & 0.251 & 0.250 & 0.250 & 0.500 & 0.253 & 0.249 \\ 0.254 & 0.253 & 0.251 & 0.251 & 0.252 & 0.252 & 0.500 & 0.249 \\ 0.249 & 0.249 & 0.249 & 0.249 & 0.249 & 0.249 & 0.249 & 0.499 \end{bmatrix} \quad (6.8b)$$

$$\Phi = \begin{bmatrix} 0.494 & 0.467 & 0.354 & 0.296 & 0.271 & 0.259 & 0.253 & 0.250 \\ 0.467 & 0.436 & 0.331 & 0.282 & 0.264 & 0.256 & 0.250 & 0.250 \\ 0.354 & 0.331 & 0.500 & 0.251 & 0.251 & 0.250 & 0.250 & 0.250 \\ 0.296 & 0.282 & 0.251 & 0.500 & 0.250 & 0.250 & 0.250 & 0.250 \\ 0.271 & 0.264 & 0.251 & 0.250 & 0.500 & 0.250 & 0.250 & 0.250 \\ 0.259 & 0.256 & 0.250 & 0.250 & 0.250 & 0.500 & 0.250 & 0.250 \\ 0.253 & 0.250 & 0.250 & 0.250 & 0.250 & 0.250 & 0.500 & 0.250 \\ 0.250 & 0.250 & 0.250 & 0.250 & 0.250 & 0.250 & 0.250 & 0.500 \end{bmatrix} \quad (6.8c)$$

Auf den Diagonalen der Matrizen (6.8b) und (6.8c) ergeben sich nicht an allen Stellen exakt die Werte 0.5; diese Abweichungen sind in der endlichen Wortlänge w begründet. Nimmt man nämlich für $P\{x = -1\} \approx 0$ an, was nach Bild 6.2 erfüllt ist, so gilt z.B. für

$$\varphi_{00} = \sum_{\lambda=1}^{2^{w-1}} P_{-\lambda} = \frac{1-p_0}{2}$$

bei symmetrischer Verteilung, so daß sich mit*) $p_0 = 0.0117$ der Wert $\varphi_{00} = 0.4941$ ergibt.

*) $p_0 = p_x(\beta=0) \cdot \Delta\beta = 3 \cdot \frac{1}{256} = 0.0117$ (vgl. Bild 6.2)

Man bemerkt ferner, daß weiter voneinander entfernte Bits oder auch benachbarte kleiner Wertigkeit sich dem Wert $1/4$ annähern. Sie werden also zunehmend unabhängiger, was unmittelbar einleuchtet.

6.4 Berechnung der Gesamtfehler-Leistung

Man geht von Gl. (4.49) für die Gesamtleistung des Fehlers aus, die allgemein ohne irgendwelche Bedingungen an die statistischen Eigenschaften der Eingangssignale gilt, solange diese unkorreliert sind.

$$N_{\Delta y} = \sum_{\rho=1}^r \alpha_{\rho}^2 \langle x_{\rho}^2(k) \rangle + 2 Q_R \sum_{\rho=1}^r \sum_{n=0}^{2^r-1} \alpha_{\rho} e_n \langle x_{\rho}(k) \cdot B(n, \mathbf{x}(k)) \rangle + \\ + Q_R^2 \sum_{n=0}^{2^r-1} \sum_{m=0}^{2^r-1} e_n e_m \langle B(n, \mathbf{x}(k)) \cdot B(m, \mathbf{x}(k)) \rangle . \quad (6.10)$$

Sie beschreibt den Fall der Zweierkomplement-Codierung ebenso wie die modifizierte Form (unter Vernachlässigung des Gleichanteils; vgl. (4.39a)), wenn für den Fehlervektor entsprechend $\mathbf{e}^{(0)}$ oder $\mathbf{e}^{(1)}$ eingesetzt wird.

Die veränderten Korrelationen zwischen den einzelnen Bits gehen dabei in die Ausdrücke

$$\langle x_{\rho}(k) \cdot B(n, \mathbf{x}(k)) \rangle \quad \text{und} \quad \langle B(n, \mathbf{x}(k)) \cdot B(m, \mathbf{x}(k)) \rangle$$

ein, die in den Anhängen 8 und 9 abgeleitet werden.

Zu ihrer Formulierung werden die Elemente der oben eingeführten Bit-Korrelations-Matrix Φ benutzt und außerdem eine Größe

$$d_{nm} = \sum_{\rho=1}^r v_{\rho} \oplus \mu_{\rho} , \quad (6.11)$$

die angibt, wie viele jeweils gleichwertige Bits der binär dargestellten Indizes n und m voneinander verschieden sind; also z.B.

$$\begin{matrix} n = 25 = & 1 & \begin{bmatrix} 1 \\ 0 \end{bmatrix} & \begin{bmatrix} 0 \\ 1 \end{bmatrix} & \begin{matrix} 0 & 1 \\ 0 & 1 \end{matrix} \end{matrix} \left. \vphantom{\begin{matrix} n = 25 = \\ m = 21 = \end{matrix}} \right\} \rightarrow d_{25,21} = 2.$$

$$m = 21 = \quad 1 \quad \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Damit erhält man nach einiger Rechnung die Ausdrücke

$$\langle x_p(k) \cdot B(n, \mathbf{x}(k)) \rangle = -(-1)^{\left[\frac{2n}{2^p} \right]} 2^{-r+1} \sigma_{x_f}^2 \quad (6.12)$$

(Anhang 8) und

$$\begin{aligned} \langle B(n, \mathbf{x}(k)) \cdot B(m, \mathbf{x}(k)) \rangle = \\ = \delta_{nm} \cdot 2^{-r} - 2 \sum_{i=1}^{w-1} 2^{-i} \varphi_{oi}^r \left(\frac{1}{2\varphi_{oi}} - 1 \right) d_{nm} + \\ + \sum_{i=1}^{w-1} \sum_{j=1}^{w-1} 2^{-i-j} \varphi_{ij}^r \left(\frac{1}{2\varphi_{ij}} - 1 \right) d_{nm} \end{aligned} \quad (6.13)$$

(Anhang 9).

Zur Gl. (6.13) ist zu bemerken, daß auftretende unbestimmte Ausdrücke der Form $(0)^0$ als 1 definiert sind.

Die Gleichungen (6.12 und 6.13) werden in (6.10) eingesetzt und ergeben die analytische Formulierung der Fehlerleistung bei beliebig verteilten, unkorrelierten Eingangssignalen:

$$\begin{aligned} N_{\Delta y} = \sigma_{x_f}^2 \sum_{p=1}^r \left(\alpha_p^2 - Q_R \alpha_p \cdot 2^{-r+2} \sum_{n=0}^{2^r-1} (-1)^{\left[\frac{2n}{2^p} \right]} e_n \right) + \\ + 2^{-r} Q_R^2 \sum_{n=0}^{2^r-1} e_n^2 + Q_R^2 \sum_{n=0}^{2^r-1} \sum_{m=0}^{2^r-1} e_n e_m \left(\sum_{i=1}^{w-1} \sum_{j=1}^{w-1} 2^{-i-j} \cdot \right. \\ \left. \cdot \varphi_{ij}^r \left(\frac{1}{2\varphi_{ij}} - 1 \right) d_{nm} - 2 \cdot \sum_{i=1}^{w-1} 2^{-i} \varphi_{oi}^r \left(\frac{1}{2\varphi_{oi}} - 1 \right) d_{nm} \right) \end{aligned} \quad (6.14)$$

mit $(A)^0 \triangleq 1 \quad \forall A \in \mathbb{R}$

6.5 Rest-Koeffizientenfehler

Wegen der Unabhängigkeit der Eingangsgrößen $x_p(k)$ kann die Definition der Rest-Koeffizientenfehler aus Kap. 4.5 hier übernommen werden:

$$\alpha_p^* = \frac{1}{\sigma_{x_f}^2} \cdot \langle \Delta y \cdot x_p \rangle. \quad (6.15)$$

Mit Hilfe der Gl. (4.32) gilt für den Mittelwert

$$\langle \Delta y \cdot x_p \rangle = \alpha_p \cdot \sigma_{x_f}^2 + Q_R \sum_{n=0}^{2^r-1} e_n \langle x_p(k) \cdot B(n, \mathbf{x}(k)) \rangle. \quad (6.16)$$

Er kann sofort hingeschrieben werden, da der Ausdruck $\langle x_p(k) \cdot B(n, \mathbf{x}(k)) \rangle$ für beliebig verteilte Eingangssignale im letzten Abschnitt bereits berechnet wurde (Gl. 6.12):

$$\langle \Delta y \cdot x_p \rangle = \alpha_p \sigma_{x_f}^2 - Q_R \cdot \sigma_{x_f}^2 \cdot 2^{-r+1} \sum_{n=0}^{2^r-1} (-1)^{\left[\frac{2n}{2^p} \right]} \cdot e_n, \quad (6.17)$$

so daß sich mit Gl. (6.15) für die Restkoeffizientenfehler ergibt

$$\alpha_p^* = \alpha_p - Q_R \cdot 2^{-r+1} \sum_{n=0}^{2^r-1} (-1)^{\left[\frac{2n}{2^p} \right]} e_n. \quad (6.18)$$

Im Vergleich mit Gl. (4.66) zeigt sich, daß die Restkoeffizientenfehler nicht von den statistischen Eigenschaften der Eingangssignale abhängen, solange diese unkorreliert sind. Dieses Ergebnis zeigte sich bereits bei der Simulation einzelner Beispiele in Kap. 6.1.

Die Leistung des linear nicht korrelierten Anteils im Gesamtfehler läßt sich unmittelbar mit Hilfe der Gl. (6.14) und (6.18) angeben:

$$N_c = N_{\Delta y} - \sigma_{x_f}^2 \sum_{p=1}^r (\alpha_p^*)^2 \quad (6.19)$$

6.6 Statistisch unterschiedliche Eingangssignale

Die Analyse der vorangegangenen Kapitel beschränkte sich auf statistisch gleichartige Eingangssignale, also mit gleichen Leistungsspektren und gleichen Amplitudenverteilungen.

Der Vollständigkeit halber soll hier gezeigt werden, wie die abgeleiteten Formeln sehr einfach auf den allgemeineren Fall statistisch unterschiedlicher (aber unkorrelierter) Eingangssignale erweitert werden können.

Dabei ändert sich an den Rest-Koeffizientenfehlern nichts: Wie oben gezeigt wurde, läßt sich aus den Ausdrücken $\langle \Delta y x_\rho \rangle$ jeweils ein Faktor $\sigma_{x_\rho}^2$, also die Leistung des Signals x_ρ , herausziehen, so daß die Rest-Koeffizientenfehler $\alpha_\rho^* = \langle \Delta y x_\rho \rangle / \sigma_{x_\rho}^2$ von den statistischen Eigenschaften der Eingangssignale vollständig unabhängig sind, auch wenn diese sich unterscheiden.

Die Gleichung (6.14) für die Fehlerleistung hingegen ist nun statt durch die φ_{ij} durch indizierte Größen $\varphi_{ij}^{(\rho)}$ zu beschreiben, da man r verschiedene Bit-Korrelations-Matrizen $\phi^{(1)} \dots \phi^{(r)}$ zu unterscheiden hat.

Dabei ergeben sich folgende Änderungen:

Die Fallunterscheidung in Gl. (A9.3) ist formal durch die Einführung der Antivalenz-Verknüpfung der Binärvariablen v_ρ und μ_ρ zu ersetzen.

$$\begin{aligned} \langle (\xi_{\rho 0}(k) \oplus v_\rho) (\xi_{\rho 1}(k) \oplus \mu_\rho) \rangle &= \begin{cases} \varphi_{oi}^{(\rho)} & \text{für } v_\rho = \mu_\rho \\ \frac{1}{2} - \varphi_{oi}^{(\rho)} & \text{für } v_\rho = \bar{\mu}_\rho \end{cases} = \\ &= \frac{1}{2} \cdot (v_\rho \oplus \mu_\rho) + (-1)^{v_\rho \oplus \mu_\rho} \cdot \varphi_{oi}^{(\rho)} \end{aligned} \quad (6.20)$$

Entsprechendes gilt für den Term ③ und ④ im Anhang 9. Nach Ersetzen der Ausdrücke

$$\varphi_{ij}^r \left(\frac{1}{2\varphi_{ij}} - 1 \right) d_{nm} + \frac{r}{\pi} \left(\frac{1}{2} (v_p \oplus u_p) + (-1)^{v_p \oplus u_p} \varphi_{ij}(\rho) \right)$$

in Gl. (6.14) erhält man die gesuchte Formulierung bei statistisch unterschiedlichen Eingangssignalen.

6.7 Analyse einzelner Beispiele

Die in den Kapiteln 6.4 und 6.5 abgeleiteten Formeln sollen auf drei Beispiele angewendet werden, die schon in Kap. 6.1 durch Simulation untersucht wurden. Zur Berechnung der gesamten Fehlerleistung dient die Gl. (6.14), für die unkorrelierte Leistung (6.19). Die Rest-Koeffizientenfehler werden hier nicht aufgeführt, weil sie gemäß Gl. (6.18) auf die gleichen Werte führen, wie bei der Analyse für weißes, gleichverteiltes Eingangssrauschen (vgl. Kap. 4.7.1).

Für die zur Berechnung benötigte Bit-Korrelations-Matrix wurde die Matrix nach Gl. (6.8b) eingesetzt.

Analyse-Ergebnisse

<u>Beispiel 1:</u>	$N_{\Delta Y_0} = 0.03964 Q_R^2$	} Zweierkomplement
	$N_{\epsilon_0} = 0.02905 Q_R^2$	
	$N_{\Delta Y_1} = 0.01023 Q_R^2$	} Modifizierte Form
	$N_{\epsilon_0} = 0.0$	

<u>Beispiel 2:</u>	$N_{\Delta Y_0} = 0.00954 Q_R^2$	} Zweierkomplement
	$N_{\epsilon_0} = 0.0$	
	$N_{\Delta Y_1} = 0.00760$	} Modifizierte Form
	$N_{\epsilon_0} = 0.0$	

<u>Beispiel_3:</u>	$N_{\Delta Y_0} = 0.05168 Q_R^2$	}	Zweier- komplement
	$N_{\epsilon_0} = 0.05131 Q_R^2$		
	$N_{\Delta Y_1} = 0.01494 Q_R^2$	}	Modifizierte Form
	$N_{\epsilon_1} = 0.01486 Q_R^2$		

Der Vergleich mit den Simulationsergebnissen in Kap. 6.1 bestätigt die Richtigkeit der Fehleranalyse für gefärbte Eingangssignale.

6.8 Ensemble-Untersuchungen

Entsprechend den Untersuchungen in Abschnitt 5.3.1 werden Ensemble-Mittelungen durchgeführt, wobei die Gesamtfehler-Leistungen jeweils nach Gl. (6.14), die unkorrelierten Anteile nach Gl. (6.19) berechnet werden. Dabei wurde das in Kap. 6.1 gewählte Beispiel für die Spektralformung der Eingangssignale angenommen, d.h. es wurde jeweils die Bit-Korrelations-Matrix (6.8b) bei der Berechnung verwendet.

Bild 6.3 zeigt die Erwartungswerte von $N_{\Delta Y}$ und N_{ϵ} bei gewöhnlichem Zweierkomplement und modifizierter Form für verschiedene Werte von r .

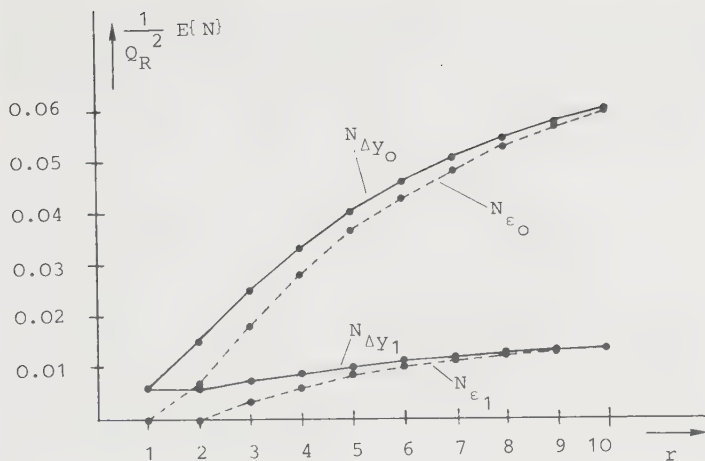


Bild 6.3: Erwartungswerte der Fehlerleistungen
(Bit-Korrelations-Matrix nach Gl. (6.8b))

Im Vergleich zu den Ergebnissen bei Erregung mit weißem, gleichverteiltem Rauschen (siehe Bild 5.5) fallen folgende Unterschiede ins Auge:

Die Fehlerleistungen sind bei farbigem Eingangssignal grundsätzlich kleiner, was mit der Reduktion der Eingangsleistungen im Einklang steht. Zum anderen näherten sich in Bild 5.5 die Größen $E\{N_{\Delta Y_0}\}$ und $E\{N_{\epsilon_0}\}$ ab etwa $r=8$ dem konstanten Wert $Q_R^2/9$ an, während hier ein solches Einlaufen innerhalb des betrachteten Wertebereiches $r=1, \dots, 10$ nicht zu beobachten ist. Untersuchungen für größere Werte r wurden nicht durchgeführt - einmal wegen erheblichem Rechenzeitaufwand bei der Auswertung der Gleichungen (6.14, 6.19) für große r , zum anderen weil die Zusammenfassung einer größeren Zahl von Variablen an einer Speicheradresse aus technologischen Gründen ohnehin nicht mehr realistisch ist.

Bei der modifizierten Form ist der Erwartungswert der gesamten Fehlerleistung nicht mehr konstant wie bei weißem, gleichverteiltem Eingangsrauschen. Lediglich die beiden Fälle $r=1$ und $r=2$ weisen gleiche Werte auf, und zwar $(\sigma_{x_f}^2 / \sigma_{x_w}^2) \cdot \frac{Q_R^2}{36}$, also

den um das Verhältnis der beiden Eingangsleistungen korrigierten Wert nach Bild 5.5. Wegen der Linearität dieser beiden Fälle ist das unmittelbar klar. Für $r > 2$ steigt $E\{N_{\Delta Y}\}$ leicht an. Ein Einlaufen auf einen Endwert ist nach Bild 6.3 nicht erkennbar.

In Bild 5.5 war die Annäherung der Leistungs-Erwartungswerte mit einem vollständigen Zurückgehen linearer Verzerrungen zugunsten eines unkorrelierten Fehleranteils verbunden, oder anders ausgedrückt: Nachdem der Gesamtfehler praktisch ausschließlich aus unkorreliertem Rauschen besteht, ändert sich die Leistung dieses Rauschens auch nicht mehr bei weiterer Steigerung der Variablen-Anzahl r . Beim farbigen Eingangssignal trifft dieses nicht zu: Bild 6.4 zeigt, daß für jedes r der prozentuale unkorrelierte Anteil am Gesamtfehler höher liegt als bei weißen, gleichverteilten Eingangssignalen.

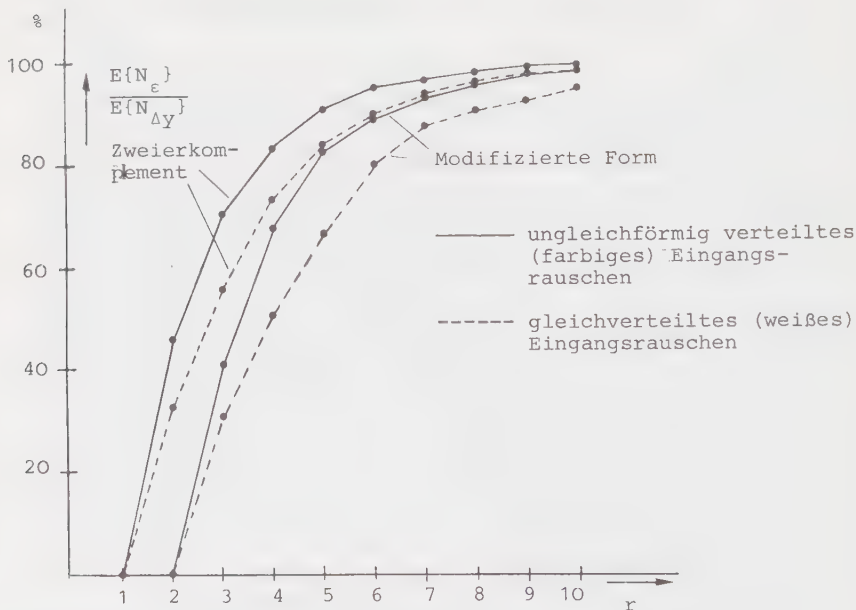


Bild 6.4: Prozentualer Anteil von unkorreliertem Rauschen am Gesamtfehler

So besteht z.B. im Zweierkomplement bei $r = 8$ der Gesamtfehler zu 98 % aus einem unkorrelierten Anteil; die durchschnittliche Fehlerleistung steigt aber für $r > 8$ weiter merklich an (siehe Bild 6.3), das heißt also: auch in den Fällen, bei denen zur hinreichend genauen Fehlerbeschreibung ausschließlich eine Rauschquelle am Ausgang angenommen werden kann, besitzt diese - anders als bei weißen, gleichverteilten Eingangssignalen - eine von r abhängige Leistung.

Um Systeme mit größeren Vektorlängen beurteilen zu können, bei denen das Gesamtergebnis aus Teilprodukten mit Vektoren jeweils der Länge r zusammengesetzt wird, trägt man entsprechend zu Kap. 5.3.1 die auf r bezogenen Erwartungswerte der Fehlerleistungen auf (Bild 6.5).

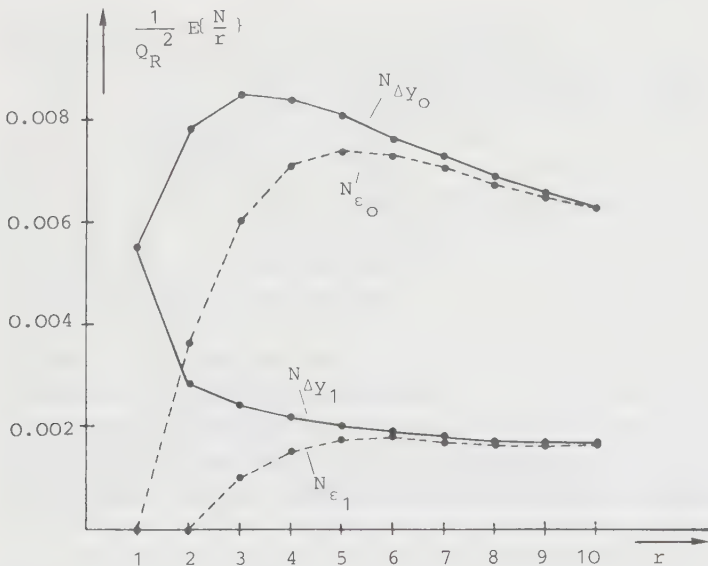


Bild 6.5: Erwartungswerte der auf r bezogenen Fehlerleistungen

Bei der Zweierkomplement-Codierung zeigt sich, daß im Gegensatz zur Erregung mit weißem Rauschen (vgl. Bild 5.7) die Konzentrierte Arithmetik ($r=1$) gegenüber allen Konstellationen

der Verteilten Arithmetik ($r=2, \dots, 10$) zu geringerer Gesamtfehlerleistung führt. Für die modifizierte Form gilt dieses zwar nicht: mit steigendem r fällt die bezogene Fehlerleistung monoton, jedoch in geringerem Maße als im Bild 5.7. Die Maxima in den Verläufen von $\frac{1}{r} E\{N_{\varepsilon_0}\}$ und $\frac{1}{r} E\{N_{\varepsilon_1}\}$ verschieben sich gegenüber denen in Bild 5.7 zu einem jeweils um 1 höheren Wert r ($r=5$ bei Zweierkomplement, $r=6$ bei modifizierter Form).

Vergleicht man also die Erregung mit gleichverteiltem und mit nicht gleichverteiltem Rauschen, so ist im letzteren Fall für die modifizierte Form eine Verringerung des Gewinns, beim Zweierkomplement sogar eine Verschlechterung im Fehlerverhalten gegenüber der Konzentrierten Arithmetik zu verzeichnen.

Diese Aussage wird auch durch die in Bild 6.6 aufgetragenen Verhältnisse der Gesamtfehler-Leistungen bei ungleichförmig und gleichförmig verteilten Eingangssignalen verdeutlicht: Da diese Größen zusätzlich auf das Eingangsleistungs-Verhältnis bezogen sind, ergibt sich für die linearen Fälle der Wert 1; mit wachsendem r steigen beide Kurven an, was ein mit zunehmender Vektorlänge ungünstigeres Fehlerverhalten bei nicht mehr gleichverteilten Eingangssignalen bedeutet. Es soll nochmals betont werden, daß mit dieser Aussage nur die Verbesserung gegenüber der Konzentrierten Arithmetik gemeint ist, nicht der Vergleich der absoluten Fehlerwerte bei den verschiedenen Erregungen. Diese liegen bei weißen, gleichverteilten Variablen wegen der größeren Variablenleistung erheblich über denen bei den hier verwendeten Signalen (Normierung in Bild 6.6 auf das Variablenleistungs-Verhältnis!).

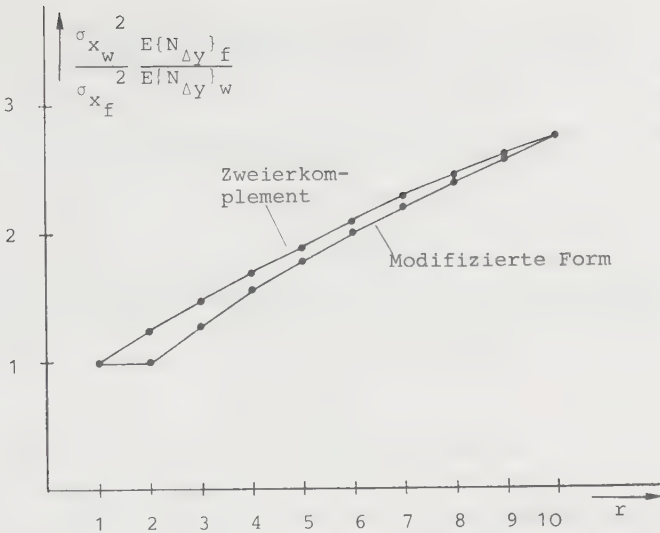


Bild 6.6: Normierte Verhältnisse der durchschnittlichen Fehlerleistungen bei gleichverteilten (farbigen) und ungleichförmig verteilten (weißen) Eingangssignalen

6.9 Abschließende Bemerkungen

In den vorangegangenen Kapiteln wurde eine geschlossene Fehleranalyse der Verteilten Arithmetik bei Erregung durch untereinander unkorrelierte Variable mit beliebigen statistischen Eigenschaften entwickelt. Dabei ging man zunächst von einer speziellen, willkürlich gewählten Signalform aus, die durch Filterung von weißen Rauschsignalen erzeugt wurde. Schon bei den Simulationsbeispielen in Kap. 6.1 zeigten sich deutliche Abweichungen in den Fehlerleistungen gegenüber der Erregung mit weißem, gleichverteiltem Rauschen, während die Rest-Koeffizientenfehler unverändert blieben.

Im weiteren wurde deutlich, daß bei Annahme von statistisch beliebigen, aber unkorrelierten Variablen nicht die Spektraleigenschaften, sondern ausschließlich die Amplitudenverteilung

gen, genauer gesagt, die sich daraus ergebenden Korrelationen zwischen verschiedenen Bit-Paaren je einer Variablen das Fehlerverhalten beeinflussen. Damit sind beispielsweise Systeme mit weißen, gaußverteilten Eingangssignalen ebenso nach den hier abgeleiteten Formeln zu behandeln wie solche mit gefärbtem Eingangsrauschen, das keine Gleichverteilung aufweist.

Zur analytischen Formulierung wird die Bit-Korrelations-Matrix (bzw. mehrere Matrizen, falls nicht alle Variablen statistisch gleichartig sind) benötigt, die sich direkt aus der Amplitudenverteilung der Variablen berechnen läßt.

7. FEHLERUNTERSUCHUNGEN BEI BERÜCKSICHTIGUNG VON KORRELATIONEN ZWISCHEN DEN VARIABLEN

7.1 Variablen-Korrelationen bei digitalen Filtern

Bei den bisherigen Untersuchungen ging man davon aus, daß die Variablen von r einzelnen Rauschquellen geliefert werden, die zwar (nach Kap. 6) beliebige Amplitudenverteilungen und Färbungen aufweisen können, aber untereinander unkorreliert sein müssen.

Man betrachte nun einen wichtigen Anwendungsfall der Verteilten Arithmetik, nämlich die Realisierung digitaler Filter. Hier trifft dieses Modell offenbar nicht zu: So hat man z.B. bei nichtrekursiven Filtern statt der r Rauschquellen nur eine einzige Signalquelle; die r verschiedenen Variablen sind dabei zeitlich zueinander verzögerte Ausgangswerte dieser Quelle.

Solange die Signalquelle weißes Rauschen liefert, bleiben die Fehleranalysen nach Kap. 4 (bei Gleichverteilung) bzw. Kap. 6 (bei beliebiger Verteilung) uneingeschränkt gültig, denn für die Autokorrelationsfunktion einer weißen Rauschgröße gilt

$$\psi_{xx}(\kappa) = \langle x(k) \cdot x(k+\kappa) \rangle = \begin{cases} \sigma_x^2 & \text{für } \kappa = 0 \\ 0 & \text{für } \kappa \neq 0 \end{cases} \quad (7.1a)$$

Sind nun die verschiedenen Eingangsvariablen Abtastwerte ein und derselben Rauschquelle zu verschiedenen Zeitpunkten

$$x_\rho(k) = x(k-\kappa_1) ; \quad x_\sigma(k) = x(k-\kappa_2) \quad (7.1b)$$

so sind sie wegen (7.1a) unkorreliert,

$$\langle x_\rho(k) \cdot x_\sigma(k) \rangle = \begin{cases} \sigma_x^2 & \text{für } \rho = \sigma \\ 0 & \text{für } \rho \neq \sigma \end{cases} \quad (7.1c)$$

und die Voraussetzungen für die Analysen nach Kap. 4 und Kap. 6 sind erfüllt.

Das ändert sich im Falle von farbigem Eingangsrauschen. Die Autokorrelationsfunktion (Bild 7.1 zeigt sie für das in Kap. 6 benutzte Beispiel von Spektralformung verschwindet jetzt für $\kappa \neq 0$ nicht mehr, so daß mit

$$\langle x_0(k) \cdot x_0(k) \rangle = \langle x(k-\kappa_1) \cdot x(k-\kappa_2) \rangle = \varphi_{xx}(\kappa_1 - \kappa_2) \quad (7.2)$$

die bisherigen Voraussetzungen verletzt werden.

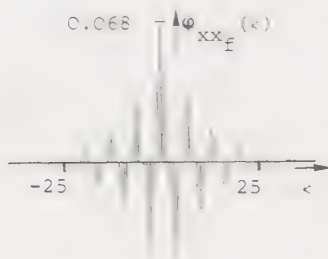


Bild 7.1: Autokorrelationsfunktion der farbigen Rauschvariablen nach Kap. 6

Für das Fehlerverhalten nichtrekursiver Filter ist nun entscheidend, in welcher Weise die Variablen gruppiert werden (in Kap. 3.2 wurde die Notwendigkeit einer Summen-Aufteilung aus Speicherplatzgründen erläutert). Dabei erhält man im Falle der Multiplex-Überlagerung der Teilsummen nach Bild 3.5b maximale zeitliche Abstände zwischen den jeweils zusammengefaßten Abtastwerten des Eingangssignals, also z.B.

$$\begin{aligned} x_1 &= x(k) \\ x_2 &= x(k-R/r) \\ x_3 &= x(k-2 R/r) \\ &\vdots \\ &\vdots \\ \text{usw.} \end{aligned}$$

Im allgemeinen nimmt die Einhüllende der Funktion $\varphi_{xx}(\kappa)$ mit

steigendem κ ab, so daß oftmals für hinreichend große Werte R/r die Korrelationen zwischen jeweils gemeinsam verarbeiteten Variablen zu vernachlässigen sind:

$$\begin{aligned} \langle x_{\rho}(k) x_{\sigma}(k) \rangle &= \langle x(k-n_1 \frac{R}{r}) \cdot x(k-n_2 \frac{R}{r}) \rangle = \varphi_{xx}((n_1-n_2) \frac{R}{r}) = \\ &= \begin{cases} \sigma_x^2 & \text{für } \rho = \sigma \\ \varphi_{xx}(\Delta n \frac{R}{r}) \approx 0 & \text{für } \rho \neq \sigma, \quad \Delta n \in \{1, \dots, r-1\} \end{cases} \quad (7.3) \end{aligned}$$

und damit die Analyse nach Kap. 6 das Fehlerverhalten näherungsweise beschreibt.

Das gilt für die parallele Realisierung nichtrekursiver Filter nach Bild 3.5a ebenfalls, wenn die frei wählbaren Gruppierungen der Variablen in gleicher Weise vorgenommen werden wie gerade beschrieben, also mit möglichst großem zeitlichen Abstand der Eingangssignal-Abtastwerte. Andernfalls müssen die Korrelationen zwischen den jeweils an einer ROM-Adresse zusammengefaßten Variablen berücksichtigt werden.

Festzuhalten bleibt also zur Fehleranalyse nichtrekursiver Filter, daß die bisher abgeleiteten analytischen Modelle in drei Fällen verwendbar sind:

- a) Erregung des Filters mit gleichverteiltem, weißem Rauschen: → Fehleranalyse exakt nach Kap. 4
- b) Erregung mit weißem, beliebig verteiltem Rauschen: → Fehleranalyse exakt nach Kap. 6
- c) Erregung mit farbigem Rauschen:
→ Fehleranalyse näherungsweise nach Kap. 6, falls bei Zusammenfassung von Abtastwerten $x(k)$ im zeitlichen Abstand $\Delta n \frac{R}{r}$ für die Autokorrelationsfunktion gilt:

$$\varphi_{xx}(\Delta n \frac{R}{r}) \approx 0 .$$

Im Gegensatz hierzu stößt man bei der Behandlung rekursiver Filter generell auf Schwierigkeiten: Erregt man das System mit weißem Rauschen, so ergeben sich im Inneren dennoch korrelierte Größen, denn die rückgeführten Signale werden ja durch das Filter selbst gefärbt. Dieses gilt für alle in Kap. 3.1 angegebenen Strukturen. Auch der bei nichtrekursiven Filtern näherungsweise behandelbare Fall c ist hier nicht denkbar, da die gemeinsam zu verarbeitenden Variablengruppen in allen Strukturen unmittelbar benachbarte Abtastwerte enthalten.

Zur Behandlung des Fehlerverhaltens rekursiver Filter ist also ein Modell erforderlich, in dem Korrelationen zwischen den einzelnen Variablen berücksichtigt werden. Überlegungen hierzu werden im folgenden durchgeführt.

7.2 Überlegungen zur Fehleranalyse bei korrelierten Variablen

7.2.1 Gesamte Fehlerleistung

Man geht von der Formulierung des Ausgangsfehlers auf der Basis der Fehlervektoren aus. Wird in Gl. (4.39a) der konstante Wert $2^{-(w-1)} Q_R \cdot e^{(1)}$ wiederum vernachlässigt, so erhält man nach Kap. 4.3 für 2^{r-1} gewöhnliches und modifiziertes Zweierkomplement die gleiche Form für den Ausgangsfehler:

$$\Delta y = \sum_{\rho=1}^r \alpha_{\rho} x_{\rho} + Q_R \sum_{n=0}^{2^r-1} e_n B(n, \mathbf{x}) \quad (7.4)$$

und damit die gesamte Leistung des Fehlers infolge endlicher Tafelwortlänge

$$\begin{aligned} N_{\Delta Y} = \langle \Delta Y^2 \rangle &= \sum_{\rho=1}^r \sum_{\sigma=1}^r \alpha_{\rho} \alpha_{\sigma} \langle x_{\rho} x_{\sigma} \rangle + \\ &+ 2 Q_R \sum_{\rho=1}^r \sum_{n=0}^{2^r-1} \alpha_{\rho} e_n \langle x_{\rho} B(n, \mathbf{x}) \rangle + \\ &+ Q_R^2 \sum_{n=0}^{2^r-1} \sum_{m=0}^{2^r-1} e_n e_m \langle B(n, \mathbf{x}) \cdot B(m, \mathbf{x}) \rangle \end{aligned} \quad (7.5)$$

Im Unterschied zu Gl. (4.49) ist zunächst zu bemerken, daß wegen der Eingangssignal-Korrelation der erste Summenausdruck sich hier nicht vereinfacht: Da für $\rho \neq \sigma$ die Terme $\langle x_\rho x_\sigma \rangle = \psi_{x_\rho x_\sigma}(0)$, also die Werte der Kreuzkorrelierten von x_ρ und x_σ bei Null nicht verschwinden, bleibt die Doppelsumme über ρ und σ stehen.

Die Berechnung der Mittelwerte $\langle x_\rho B(n, \mathbf{x}) \rangle$ und $\langle B(n, \mathbf{x}) \cdot B(m, \mathbf{x}) \rangle$ für weißes, gleichverteiltes Eingangsrauschen (Anhang 3 und 4) ging von der Definition des Gewichtes $B(n, \mathbf{x})$ und der Binärdarstellung von x_ρ aus. Eine geschlossene Lösung dafür (wie auch in Kap. 6) gelang deshalb, weil in den Gleichungen (A3.1) bzw. (A4.1) wegen der statistischen Unabhängigkeit der Bits verschiedener Variablen Produkt- und Mittelungszeichen vertauscht werden konnten. Hier liegen nun zwischen den Variablen Korrelationen vor, womit natürlich auch die Bits ihrer Binärdarstellung nicht mehr unabhängig sind.

Ausdrücke der Form

$$\langle \prod_{\rho=1}^r \xi_{\rho i} \rangle$$

sind also nicht mehr durch das Produkt der einzelnen Mittelwerte zu beschreiben; vielmehr bedeutet gerade die Abweichung davon ein Maß für die Abhängigkeit der Bits verschiedener Variablen, über die ja vorab keine Aussage gemacht werden kann. Die Berechnung der Mittelwerte $\langle x_\rho B(n, \mathbf{x}) \rangle$ und $\langle B(n, \mathbf{x}) \cdot B(m, \mathbf{x}) \rangle$ ist also nicht möglich, ohne diese Bit-Korrelationen im einzelnen zu kennen.

Bei der Fehleranalyse geht man daher am einfachsten von der Gl. (7.5) direkt aus, in die die durch Simulation zu ermittelnden Größen

$$\begin{aligned} \psi_{x_\rho x_\sigma}(0) &= \langle x_\rho \cdot x_\sigma \rangle & \rho, \sigma \in \{1, \dots, r\} \\ \langle x_\rho \cdot B(n, \mathbf{x}) \rangle & & \rho \in \{1, \dots, r\}, n \in \{0, \dots, 2^r - 1\} \\ \langle B(n, \mathbf{x}) \cdot B(m, \mathbf{x}) \rangle & & n, m \in \{0, \dots, 2^r - 1\} \end{aligned}$$

eingesetzt werden.

Das heißt, daß eine Menge von $r^2 + r \cdot 2^r + 2^{2r}$ jeweils durch Simulation gefundenen Werte übergeben werden muß. Im Fall $r=8$ bedeutet das eine Datenmenge von 67648 Werten; er ist also bereits mit vernünftigem Aufwand nicht mehr zu bewältigen. Im Rahmen dieser Arbeit beschränkt man sich deshalb auf die Untersuchung von Systemen mit $r \leq 5$ (die Zahl der Werte ist dann ≤ 1209); der Fall eines rekursiven Blockes 2. Grades ist damit eingeschlossen.

Auf den ersten Blick befriedigt dieses Verfahren nicht. Man wird zu Recht einwenden, daß die direkte Simulation von einzelnen zu untersuchenden Systemen auf Bit-Ebene weitaus einfacher ist, sowohl vom Speicher- als auch vom Rechenzeit-Aufwand her.

Wenn das hier beschriebene Verfahren dennoch angewendet wird, dann deshalb, weil nur so Ensemble-Untersuchungen durchgeführt werden können, die in dieser Arbeit wegen der Verdeutlichung typischer Fehlereigenschaften von besonderem Interesse sind. Betrachtet man nämlich Gl. (7.5), so erkennt man die nach wie vor bestehende formale Trennung von Systemparameter- und Eingangssignal-Einflüssen. Es ist also möglich, vorab in einem einzigen Simulationsslauf diejenigen Terme zu bestimmen, die durch die Eingangssignal-Statistik geprägt werden, und damit dann für ein großes Ensemble von Koeffizientensätzen statistische Untersuchungen durchzuführen. Man erhält damit die gewünschten Aussagen über im Mittel zu erwartende Fehlereigenschaften, natürlich nur für den bei der vorangegangenen Simulation angenommenen speziellen Eingangssignal-Typ.

Auf der Basis von direkter Simulation jedes einzelnen Systems des betrachteten Ensembles wären derartige Untersuchungen ausgeschlossen.

7.2.2 Rest-Koeffizientenfehler

Auch die Bestimmung der Rest-Koeffizientenfehler verkompliziert sich hier gegenüber der Analyse bei unkorrelierten Eingangssignalen. Wie dort geht man zunächst von dem Ansatz

$$\Delta y = \sum_{\rho=1}^r \alpha_{\rho}^{*} x_{\rho} + \varepsilon \quad (7.6)$$

aus, in dem die Eingangssignale und der Fehleranteil ε nicht miteinander korreliert sein sollen. Daraus läßt sich ein lineares Gleichungssystem zur Bestimmung der Rest-Koeffizientenfehler $\alpha_1^{*}, \dots, \alpha_r^{*}$ gewinnen:

$$\begin{aligned} \langle \Delta y \cdot x_1 \rangle &= \alpha_1^{*} \langle x_1^2 \rangle + \alpha_2^{*} \langle x_1 x_2 \rangle + \dots + \alpha_r^{*} \langle x_1 x_r \rangle \\ \vdots & \qquad \qquad \qquad \vdots \\ \langle \Delta y \cdot x_r \rangle &= \alpha_1^{*} \langle x_1 x_r \rangle + \alpha_2^{*} \langle x_2 x_r \rangle + \dots + \alpha_r^{*} \langle x_r^2 \rangle \end{aligned} \quad (7.7)$$

Im Gegensatz zu Gl. (4.64) in Kap. 4.5 sind die Bestimmungsgleichungen für die α_{ρ}^{*} nunmehr verkoppelt, da die Terme $\langle x_{\rho} x_{\sigma} \rangle$ für $\rho \neq \sigma$ nicht mehr verschwinden. Bei Ensemble-Untersuchungen braucht allerdings die zur Lösung von (7.7) erforderliche Matrix-Inversion nur einmal durchgeführt zu werden, denn die Matrix besteht aus den Elementen $\langle x_{\rho} x_{\sigma} \rangle = \varphi_{x_{\rho} x_{\sigma}}(0)$, d.h. den Werten der Kreuzkorrelierten bei Null. Sie sind also bei Kenntnis der Variablen-Statistik vorab einzusetzen. Für jeden Koeffizientensatz des betrachteten Ensembles ist zur Berechnung der Rest-Koeffizientenfehler also nur eine Matrix-Multiplikation erforderlich:

$$(\mathbf{a}^{*} - \mathbf{a}) = \mathbf{Q}_R \cdot \mathbf{A}^{-1} \cdot \mathbf{b} \quad (7.8a)$$

mit den Elementen des Spaltenvektors $\mathbf{b}$

$$b_{\rho} = \sum_{n=0}^{2^{r-1}-1} e_n \cdot \langle x_{\rho} B(n, \mathbf{x}) \rangle, \quad \rho \in \{1, \dots, r\} \quad (7.8b)$$

und denen der symmetrischen Matrix $\mathbf{A}$

$$a_{\rho\sigma} = \varphi_{x_{\rho} x_{\sigma}}(0), \quad \rho, \sigma \in \{1, \dots, r\} \quad (7.8c)$$

Der Vektor **b** enthält dabei die Information über das Fehlerverhalten des individuellen Systems, ausgedrückt durch die Fehlervektor-Komponenten.

7.2.3 Unkorrelierte Fehlerleistung

Wegen der statistischen Unabhängigkeit des Fehleranteils ε von den Eingangsvariablen x_ρ gilt:

$$\begin{aligned} N_{\Delta Y} &= \left\langle \left(\sum_{\rho=1}^r \alpha_\rho^* x_\rho \right)^2 \right\rangle + N_\varepsilon = \\ &= \sum_{\rho=1}^r \sum_{\sigma=1}^r \alpha_\rho^* \alpha_\sigma^* \varphi_{x_\rho x_\sigma}(0) + N_\varepsilon \end{aligned} \quad (7.9)$$

Die Werte $\varphi_{x_\rho x_\sigma}(0)$ sind aus der Simulation zur Formulierung der statistischen Eingangssignal-Eigenschaften bekannt. Nachdem die gesamte Fehlerleistung mit Gl. (7.5) und die Rest-Koeffizientenfehler mit Gl. (7.8) bestimmt sind, kann die Leistung des unkorrelierten Fehleranteils aus Gl. (7.9) berechnet werden:

$$N_\varepsilon = N_{\Delta Y} - \sum_{\rho=1}^r \sum_{\sigma=1}^r \alpha_\rho^* \alpha_\sigma^* \varphi_{x_\rho x_\sigma}(0) \quad (7.10)$$

7.3 Untersuchung einzelner Beispiele

7.3.1 Simulation zur Formulierung der Variablen-Statistik

Wegen des wichtigen Anwendungsfalles der Verteilten Arithmetik auf digitale Filter werden für die folgenden Untersuchungen korrelierte Variablen nach Bild 7.2 erzeugt.

Dabei wird zur Spektralformung ein Filter 2. Grades mit der Übertragungsfunktion nach Gl. (6.1) eingesetzt. Die Zählerkonstante bestimmt man wieder so, daß nach 100 000 Abtast-

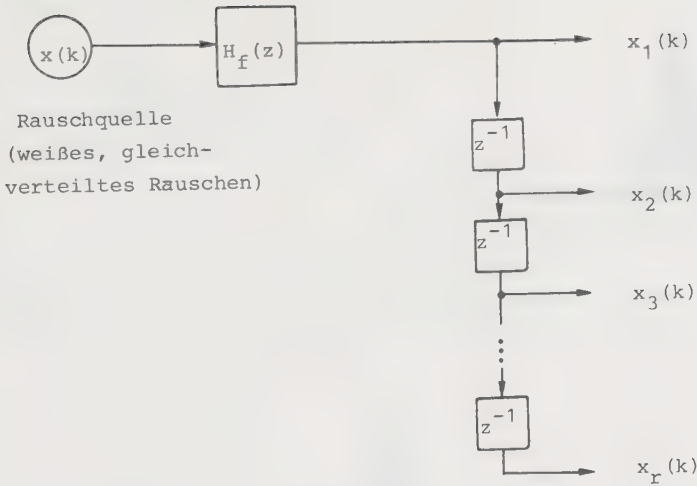


Bild 7.2: Erzeugung von r untereinander korrelierten Rauschsignalen

werten gerade keine Überläufe am Ausgang des Filters stattgefunden haben (vgl. Kap. 6.1).

Insgesamt werden drei verschiedene Signaltypen verwendet, die sich bei festem Polradius ($\rho_\infty = 0.9$) im Polwinkel ψ_∞ des spektralformenden Filters unterscheiden. Sie sind in den Bildern 7.3a-c zusammengestellt (zur Unterscheidung werden die drei Signalformen hier und im folgenden mit f_1 , f_2 und f_3 gekennzeichnet).



Bild 7.3a: Frequenzgang des Spektralformungsfilters, Autokorrelationsfunktion der Variablen $x_p(k)$ beim Signaltyp $x_{f1}(k)$ mit $\rho = 0.9$, $\psi = 0.3\pi$



Bild 7.3b: Frequenzgang des Spektralformungsfilters, Autokorrelationsfunktion der Variablen $x_{\rho}(k)$ beim Signaltyp $x_{f2}(k)$ mit $\rho_{\infty} = 0.9$, $\psi_{\infty} = 0.1\pi$

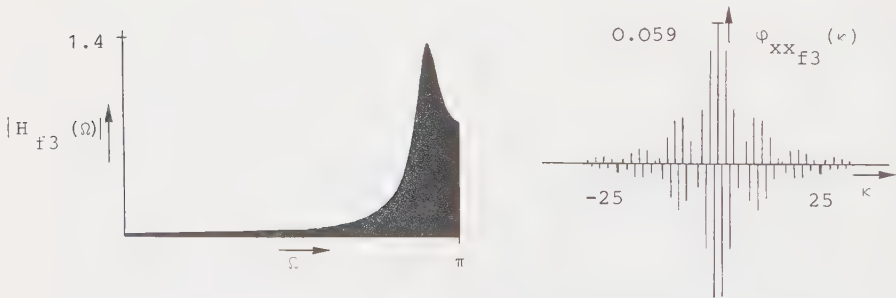


Bild 7.3c: Frequenzgang des Spektralformungsfilters, Autokorrelationsfunktion der Variablen $x_{\rho}(k)$ beim Signaltyp $x_{f3}(k)$ mit $\rho_{\infty} = 0.9$, $\psi_{\infty} = 0.9\pi$

Durch Simulation sollen nun die in Gl. (7.5), (7.8) und (7.10) benötigten zeitlichen Mittelwerte bestimmt werden, durch die der Einfluß der Variablen-Statistik auf das Fehlerverhalten individueller Systeme in Verteilter Arithmetik festgelegt ist. Bild 7.4 zeigt das prinzipielle Blockdiagramm dafür. Es sei erwähnt, daß wegen der Erzeugung der korrelierten Signale nach Bild 7.2, also durch zeitliche Verzögerung der Folge $x_f(k)$ die Kreuzkorrelierten bei Null durch Werte der Autokorrelationsfunktion ersetzt werden können:

$$\psi_{x_{\rho} x_{\sigma}}(0) = \psi_{x_f x_f}(\rho - \sigma) \quad (7.11)$$

Zur Veranschaulichung sind in Tabelle 7.1 die so gewonnenen Ausdrücke für das Beispiel $r = 2$ beim Signaltyp x_{f1} zusammen-

gestellt. Zur Mittelung wurden dabei jeweils $K = 75\,000$ Abtastwerte benutzt.

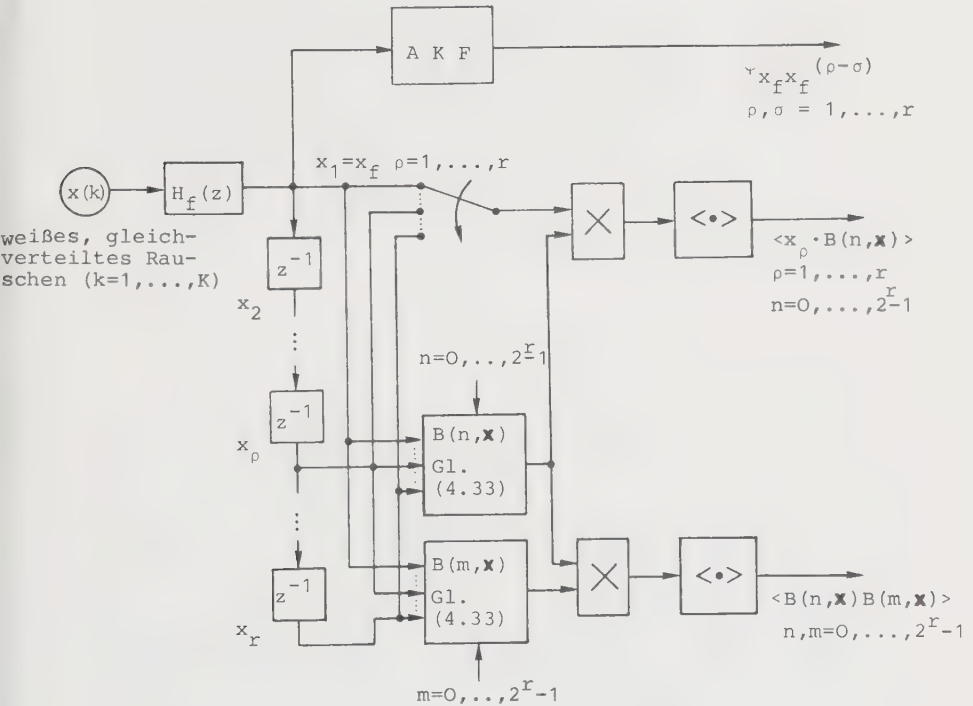


Bild 7.4: Blockdiagramm zur Berechnung der Variablen-Statistik durch Simulation

$$\phi_{x_{f1} x_{f1}}(0) = \sigma_{x_{f1}}^2 = 0.067968$$

$$\phi_{x_{f1} x_{f1}}(1) = 0.035140$$

$\langle x_\rho \cdot B(n, \mathbf{x}) \rangle :$

$n \downarrow$	$\rho \rightarrow 1$	2
0	-0.5196E-01	-0.5196E-01
1	0.1382E-01	-0.1349E-01
2	-0.1350E-01	0.1382E-01
3	0.5163E-01	0.5163E-01

(zu Tabelle 7.1)

$\langle B(n, \mathbf{x}) \cdot B(m, \mathbf{x}) \rangle :$

$n \downarrow$	$m \rightarrow$	0	1	2	3
0		0.7883E-01	-0.2659E-01	-0.2660E-01	-0.2510E-01
1		-0.2659E-01	0.3984E-01	0.1252E-01	-0.2627E-01
2		-0.2660E-01	0.1252E-01	0.3985E-01	-0.2628E-01
3		-0.2510E-01	-0.2627E-01	-0.2628E-01	0.7818E-01

Tabelle 7.1: Ergebnisse der Simulation für $r = 2$
beim Signaltyp $x_{f1}(k)$

7.3.2 Beispiele

Am Schluß dieses Abschnitts sind in den Tabellen 7.2a-f die Analyse-Ergebnisse für einige Beispiele von Koeffizientensätzen zusammengestellt. Dabei werden die Werte bei Variablen-Korrelationen denen ohne Korrelation (Analyse nach Kap. 6.3/6.4) gegenübergestellt.

Es interessiert zunächst der Fall $r = 2$: Hier führt bekanntlich die modifizierte Zweierkomplement-Codierung immer zu linearen Systemen (vgl. Kap. 4.72). Demgemäß müssen im Beispiel nach Tabelle 7.2a die Rest-Koeffizientenfehler bei korrelierten und unkorrelierten Variablen identisch sein (und sich mit denen nach Gl. (4.66) für weißes, gleichverteiltes Rauschen decken). Leichte Abweichungen sind mit Simulationsungenauigkeiten bei der Berechnung der Variablen-Statistik zu erklären. Außerdem muß der unkorrelierte Fehleranteil N_{ϵ_1} verschwinden, was im betrachteten Beispiel erfüllt ist.

In der gesamten Fehlerleistung besteht zwischen den beiden Fällen ein Unterschied, der von der nicht verschwindenden Kovarianz der korrelierten Signale herrührt (Zahlenwerte $\gamma_{x_0 x_0}^{(0)}$ vgl. Tabelle 7.1):

- unkorrelierte Variablen

$$\begin{aligned} N_{\Delta Y_1} &= (\alpha_1^*)^2 \sigma_{x_1}^2 + (\alpha_2^*)^2 \sigma_{x_2}^2 = \\ &= ((0.3826)^2 + (0.0644)^2) \cdot 0.067968 Q_R^2 \\ &= 0.010231 Q_R^2 \end{aligned}$$

- korrelierte Variable

$$\begin{aligned} N_{\Delta Y_1} &= (\alpha_1^*)^2 \sigma_{x_{f1}}^2 + (\alpha_2^*)^2 \sigma_{x_{f1}}^2 + \\ &+ 2 \cdot (\alpha_1^* \alpha_2^*) \varphi_{x_{f1} x_{f1}}^{(1)} = \\ &= 0.010231 Q_R^2 + 2 \cdot (0.3826 \cdot 0.0644) \cdot 0.03514 Q_R^2 \\ &= 0.01196 Q_R^2 \end{aligned}$$

In den weiteren drei Beispielen für $r = 2$ wird der eben benutzte Koeffizientensatz für Zweierkomplement-Codierung untersucht, und zwar für die Signaltypen f_1 , f_2 und f_3 (Tabellen 7.2b-d). Es zeigen sich in den Leistungswerten $N_{\Delta Y_0}$ und N_{ϵ_0} deutliche Unterschiede zwischen den Analyse-Ergebnissen mit und ohne Berücksichtigung der Korrelationen. Bemerkenswert hingegen ist die Unveränderlichkeit der Rest-Koeffizientenfehler (im Rahmen der Simulationsgenauigkeit), obwohl in dem betrachteten Beispiel ein nichtlineares System vorliegt. Es wurde eine Reihe von weiteren Systemen mit $r = 2$ untersucht, mit verschiedenen Koeffizientensätzen und verschiedenen Signaltypen, die jeweils gemäß Bild 7.2 erzeugt wurden unter Variation von Pol-Winkel und -Radius des spektralformen- den Filters. Dabei erwies sich in allen Fällen, daß hier die Rest-Koeffizientenfehler offenbar nicht durch die Variablen-Korrelation beeinflusst werden. Sie hängen nur von den jeweiligen Koeffizienten ab und können deshalb stets nach Gl. (4.66), also der Beziehung für weiße, gleichverteilte Variablen, berechnet werden.

Für Systeme mit mehr als zwei Koeffizienten gilt dies nicht, wie die in den Tabellen 7.2e und f aufgeführten Fälle für $r = 4$ deutlich machen. Das erste dieser beiden Beispiele (Zweierkomplement) zeigt noch etwas anderes: Bisher waren die Rest-Koeffizientenfehler grundsätzlich auf den Wertebereich der a-priori-Koeffizientenfehler, also $\left[-\frac{Q_R}{2}, \frac{Q_R}{2}\right]$ beschränkt. Hier ergibt sich nun für $\alpha_2^* = 0.61475 Q_R$; d.h. infolge von Variablen-Korrelationen können die Rest-Koeffizientenfehler offenbar den Wertebereich $\left[-\frac{Q_R}{2}, \frac{Q_R}{2}\right]$ überschreiten.

Schließlich ist noch eine weitere prinzipielle Eigenschaft an diesem letzten Beispiel zu beobachten: In Tabelle 7.2e erscheint ohne Annahme von Korrelationen der Koeffizientenfehler α_3 als Rest-Koeffizientenfehler wieder. In Kap. 4.7.2 wurde auf derartige Fälle hingewiesen: Bezüglich einzelner Variablen verhält sich das System "linear". Werden nun Korrelationen zwischen den Variablen berücksichtigt, so verliert das System diese Eigenschaft; im betrachteten Beispiel ist nämlich dann $\alpha_3^* \neq \alpha_3$. Man macht sich das sofort anhand des Gleichungssystems (7.8) klar: Liegen keine Korrelationen vor, so ergibt sich wegen

$$\varphi_{x_i x_j}(0) = 0 \quad \text{für} \quad \rho \neq \sigma$$

für die Koeffizientenmatrix eine Diagonalmatrix, die Bestimmungsgleichungen sind entkoppelt. Soll nun also für einen der Rest-Koeffizientenfehler gelten

$$\alpha_\rho^* = \alpha_\rho,$$

so muß das entsprechende Element des Vektors **b** verschwinden. Mit Gl. (6.12) muß also offenbar gelten

$$\sum_{n=0}^{2^{r-1}} (-1)^{\left\lfloor \frac{2n}{2^\rho} \right\rfloor} e_n = 0,$$

eine Bedingung, die nur vom Fehlervektor, also der Koeffizienten-Konstellation abhängt.

Verschwinden die Signalkorrelationen nicht, so ist das verkoppelte Gleichungssystem (7.8) zu lösen und α_ρ^* wird von den statistischen Eigenschaften der Variablen abhängig. Im allge-

meinen wird dann der Wert $(\alpha_{\rho}^* - \alpha_{\rho})$ auch bei Erfüllung der obigen Fehlervektor-Bedingung nicht mehr verschwinden.

Zusammenstellung der betrachteten Beispiele:

$$-\alpha_1 = -0.1174 Q_R$$

$$-\alpha_2 = -0.4356 Q_R$$

	ohne Korrelation	mit Korrelation
$\alpha_1^{*(1)}$	$0.3826 Q_R$	$0.3827 Q_R$
$\alpha_2^{*(1)}$	$0.0644 Q_R$	$0.0644 Q_R$
$N_{\Delta Y_1}$	$0.01023 Q_R^2$	$0.01174 Q_R^2$
N_{ε_1}	0.0	0.0

Tabelle 7.2a: Beispiel $r = 2$ bei modifiziertem Zweierkomplement beim Signaltyp f1 ($\rho_{\infty} = 0.9$, $\psi_{\infty} = 0.3\pi$)

$$-\alpha_1 = -0.1174 Q_R$$

$$-\alpha_2 = -0.4356 Q_R$$

	ohne Korrelation	mit Korrelation
$\alpha_1^{*(0)}$	$0.3826 Q_R$	$0.3812 Q_R$
$\alpha_2^{*(0)}$	$0.0644 Q_R$	$0.0638 Q_R$
$N_{\Delta Y_0}$	$0.03964 Q_R^2$	$0.0383 Q_R^2$
N_{ε_0}	$0.02905 Q_R^2$	$0.0264 Q_R^2$

Tabelle 7.2b: Beispiel für $r = 2$ bei gewöhnlichem Zweierkomplement beim Signaltyp f1 ($\rho_{\infty} = 0.9$, $\psi_{\infty} = 0.3\pi$)

$$\alpha_1 = -0.1174 Q_R$$

$$\alpha_2 = -0.4356 Q_R$$

	ohne Korrelation	mit Korrelation
$\alpha_1^{*(0)}$	$0.3826 Q_R$	$0.3827 Q_R$
$\alpha_2^{*(0)}$	$0.0644 Q_R$	$0.0653 Q_R$
$N_{\Delta y_O}$	$0.04358 Q_R^2$	$0.02865 Q_R^2$
N_{ε_O}	$0.03363 Q_R^2$	$0.01601 Q_R^2$

Tabelle 7.2c: Beispiel für $r = 2$ bei Zweierkomplement
Signaltyp f2 ($\rho_\infty = 0.9, \psi_\infty = 0.1\pi$)

$$\alpha_1 = -0.1174 Q_R$$

$$\alpha_2 = -0.4356 Q_R$$

	ohne Korrelation	mit Korrelation
$\alpha_1^{*(0)}$	$0.3826 Q_R$	$0.3855 Q_R$
$\alpha_2^{*(0)}$	$0.0644 Q_R$	$0.0644 Q_R$
$N_{\Delta y_O}$	$0.03398 Q_R^2$	$0.02026 Q_R^2$
N_{ε_O}	$0.02507 Q_R^2$	$0.01402 Q_R^2$

Tabelle 7.2d: Beispiel für $r = 2$ bei Zweierkomplement
Signaltyp f3 ($\rho_\infty = 0.9, \psi_\infty = 0.9\pi$)

$$-\alpha_1 = -0.3443 Q_R$$

$$-\alpha_2 = -0.1967 Q_R$$

$$-\alpha_3 = -0.0164 Q_R$$

$$-\alpha_4 = -0.3516 Q_R$$

	ohne Korrelation	mit Korrelation
$\alpha_1^{*(0)}$	$0.1557 Q_R$	$0.06809 Q_R$
$\alpha_2^{*(0)}$	$0.3032 Q_R$	$0.61475 Q_R$
$\alpha_3^{*(0)}$	$-0.0164 Q_R$	$-0.08866 Q_R$
$\alpha_4^{*(0)}$	$0.1484 Q_R$	$0.02752 Q_R$
$N_{\Delta Y_O}$	$0.04669 Q_R^2$	$0.04940 Q_R^2$
N_{ε_O}	$0.03728 Q_R^2$	$0.02396 Q_R^2$

Tabelle 7.2e: Beispiel für $r = 4$; gewöhnliches Zweierkomplement beim Signaltyp f1 ($\rho_\infty=0.9, \psi_\infty=0.3\pi$)

$$-\alpha_1 = -0.3443 Q_R$$

$$-\alpha_2 = -0.1967 Q_R$$

$$-\alpha_3 = -0.0164 Q_R$$

$$-\alpha_4 = -0.3516 Q_R$$

	ohne Korrelation	mit Korrelation
$\alpha_1^{*(1)}$	$0.03068 Q_R$	$0.071406 Q_R$
$\alpha_2^{*(1)}$	$-0.07174 Q_R$	$-0.165824 Q_R$
$\alpha_3^{*(1)}$	$0.10855 Q_R$	$0.051052 Q_R$
$\alpha_4^{*(1)}$	$0.02338 Q_R$	$0.040205 Q_R$
$N_{\Delta Y_1}$	$0.010603 Q_R^2$	$0.004759 Q_R^2$
N_{ε_1}	$0.009351 Q_R^2$	$0.003930 Q_R^2$

Tabelle 7.2f: Beispiel für $r = 4$; modifiziertes Zweierkomplement beim Signaltyp f1 ($\rho_\infty=0.9, \psi_\infty=0.3\pi$)

7.4 Ensemble-Untersuchungen

7.4.1 Erwartungswerte der Fehlerleistungen

Durch Ensemble-Mittelung werden die Erwartungswerte der Fehlerleistungen $N_{\Delta y}$ und N_{ϵ} für gewöhnliches und modifiziertes Zweierkomplement berechnet. Dabei werden für jeden individuellen Koeffizientensatz des Ensembles die Gleichungen (7.5) und (7.10) ausgewertet; die von den statistischen Eigenschaften der Variablen abhängigen Terme werden gemäß Kap. 7.3.1 durch Simulationsvorläufe bestimmt.

Bild 7.5 zeigt die Ergebnisse für den Signaltyp x_{f1} .

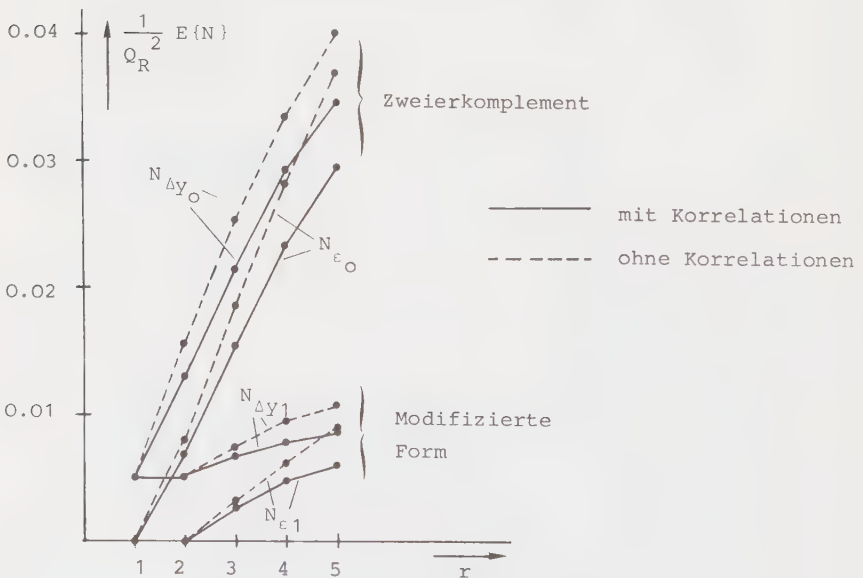


Bild 7.5: Erwartungswerte der Fehlerleistungen beim Signaltyp x_{f1} ($\rho_{\infty}=0.9, \psi_{\infty}=0.3\pi$)

Zum Vergleich sind hier gestrichelt die Verläufe der Erwartungswerte eingetragen, die sich nach der Analyse von Kap. 6, also ohne Berücksichtigung von Korrelationen ergeben. Man sieht, daß diese Werte für alle hier betrachteten r obere Schranken darstellen. Es stellt sich nun die Frage, ob das allgemein bei beliebigen Variablen-Korrelationen gilt, denn in dem Falle hätte man in der sehr viel einfacher zu handhabenden Analyse nach Kap. 6 eine brauchbare Abschätzung für das Fehlerverhalten bei korrelierten Variablen, indem diese Korrelationen einfach vernachlässigt werden.

Eine kurze Betrachtung des leicht überschaubaren Falles $r = 2$ soll helfen, gezielt eine andere Form der Variablen-Korrelation zu wählen, um diese Frage zu klären.

Für den Erwartungswert der Gesamtfehler-Leistung gilt für $r = 2$ (Zweierkomplement):

$$\begin{aligned} E\{N_{\Delta Y_0}\} &= \sigma_{x_f}^2 \left(E\{(\alpha_1^*)^2\} + E\{(\alpha_2^*)^2\} \right) + \\ &+ 2\sigma_{x_f}^2(1) \cdot E\{\alpha_1^* \alpha_2^*\} + E\{N_{\varepsilon_0}\} \end{aligned} \quad (7.12)$$

Im letzten Abschnitt wurde durch Einzelfall-Untersuchungen gefunden, daß im Fall $r = 2$ die Rest-Koeffizientenfehler durch Variablen-Korrelationen nicht beeinflußt werden. Man erhält damit als Unterschied zwischen den Fehlerleistungen mit und ohne Berücksichtigung von Korrelationen

$$\begin{aligned} E\{N_{\Delta Y_0}\} \Big|_{\text{korr}} - E\{N_{\Delta Y_0}\} \Big|_{\text{unkorr}} &= \\ &= 2\sigma_{x_f}^2(1) \cdot E\{\alpha_1^* \alpha_2^*\} + E\{N_{\varepsilon_0}\} \Big|_{\text{korr}} - E\{N_{\varepsilon_0}\} \Big|_{\text{unkorr}} \end{aligned} \quad (7.13)$$

Es ist mit Hilfe von Gl. (6.18) leicht zu zeigen, daß bei Annahme von gleichverteilten a-priori-Koeffizientenfehlern gilt

$$E\{\alpha_1^* \alpha_2^*\} = -\frac{Q_R^2}{48} \quad (7.14)$$

Sieht man zunächst von den unterschiedlichen Erwartungswerten von N_{ε_0} ab, so ergibt sich also infolge der Variablen-Korrela-

tion eine Reduktion der Gesamtfehler-Leistung um den Wert $\frac{Q_R^2}{24} \varphi_{x_f x_f}(1)$.

Der Signaltyp f_1 führt demgemäß mit $\varphi_{x_{f1} x_{f1}}(1) = 0.03514 > 0$ zu einer Fehlerleistungs-Verminderung bei Berücksichtigung von Korrelationen. Bei Signalen mit negativen Kovarianzen $\varphi_{x_f x_f}(1)$ müßte nun nach diesen Überlegungen eine entsprechende Fehlerleistungs-Steigerung erfolgen, wenn nicht die durch die Variablen-Korrelation bewirkte Veränderung von $E\{N_{\varepsilon_0}\}$ den beschriebenen Einfluß wieder aufhebt.

Die folgende Untersuchung zeigt jedoch, daß gerade dieses offenbar der Fall ist: In Bild 7.6 ist für Zweierkomplement der Verlauf der Fehlerleistungs-Erwartungswerte wiedergegeben, und zwar für den Signaltyp x_{f3} mit der negativen Kovarianz $\varphi_{x_{f3} x_{f3}}(1) = -0.05597$. Die Gesamtfehler-Leistungswerte liegen auch hier unterhalb der gestrichelten Kurve, also den Analyse Ergebnissen bei Vernachlässigung der Korrelationen. Gegenüber Bild 7.5 ist eine merkliche Verringerung des Fehleranteils N_{ε_0} festzustellen (vgl. auch Bild 7.7).

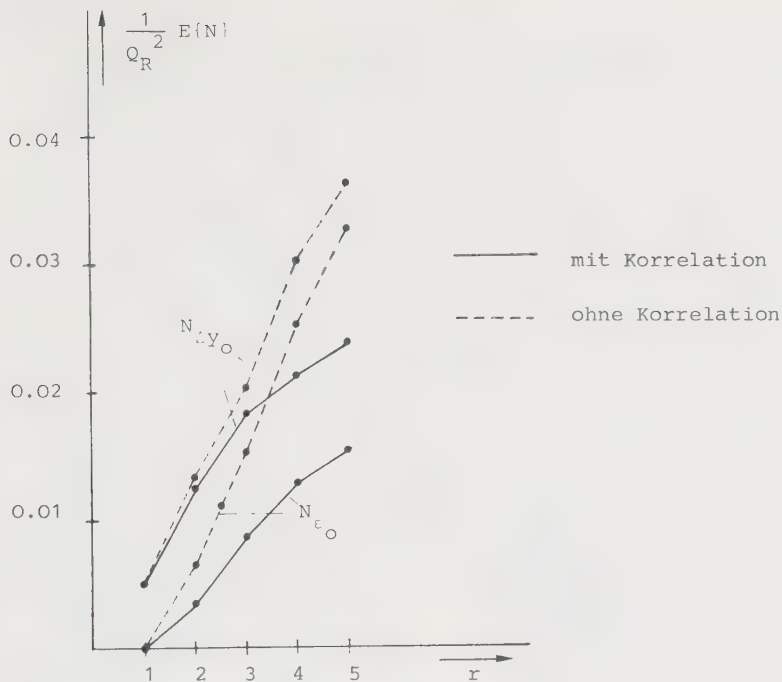


Bild 7.6: Erwartungswerte der Fehlerleistungen beim Signaltyp x_{f3} ($\rho_{\infty}=0.9, \varphi_{\infty}=0.9$) (Zweierkomplement)

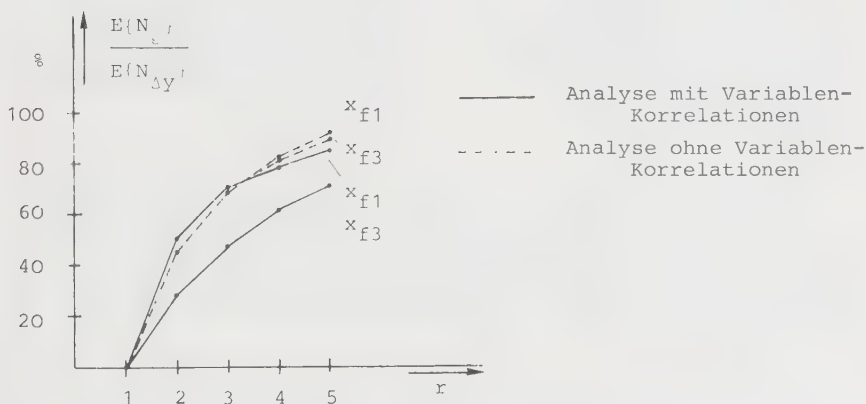


Bild 7.7: Verhältnisse von unkorrelierten Fehleranteilen zu Gesamtfehlern (Zweierkomplement)

Die untersuchten Beispiele legen den Schluß nahe, daß die Vernachlässigung von Signal-Korrelationen zu einer oberen Abschätzung der Fehlerleistung führt. Die Qualität dieser Abschätzung hängt dabei naturgemäß von den jeweiligen speziellen Eigenschaften der Variablen ab: Bei den hier betrachteten Beispielen betragen die Abweichungen bis zu 16 % für $E\{N_{\Delta y_O}\}$ bzw. 26 % für $E\{N_{\varepsilon_O}\}$.

Abschließend bleibt festzustellen, daß diese Aussagen nur im statistischen Mittel Gültigkeit haben. Im Einzelfall können sich durchaus größere Differenzen ergeben; die Beispiele in Kap. 7.3.2 machen das deutlich. D.h. bei der Untersuchung von Einzelfällen sind die Signal-Korrelationen im allgemeinen nicht vernachlässigbar: Es sei deshalb nochmals betont, daß der einfachste Weg der Analyse individueller Systeme die direkte Simulation auf Bit-Ebene ist.

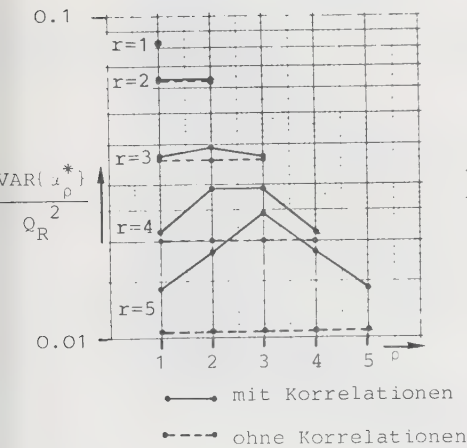
7.4.2 Rest-Koeffizientenfehler

Die Fehleruntersuchungen bei korrelierten Variablen werden mit statistischen Betrachtungen der Rest-Koeffizientenfehler abgeschlossen. Bei der Behandlung großer Ensembles von Koeffizientensätzen wird jeweils das Gleichungssystem (7.8) ausgewertet.

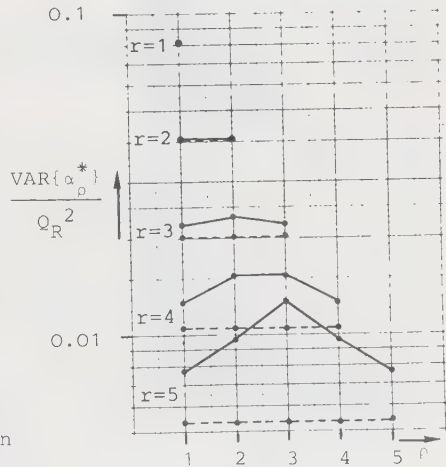
Im Gegensatz zu den Untersuchungen bei unkorrelierten Variablen wird hier bei der Berechnung der Rest-Koeffizientenfehler-Varianzen nach verschiedenen Indizes ρ unterschieden, denn je nach der Position der betrachteten Variablen innerhalb des Schiebeprozesses liegen verschiedene Korrelationen zu ihren $r-1$ benachbarten Signalen vor.

Bild 7.8a zeigt die Varianzen der Rest-Koeffizientenfehler für den Signaltyp x_{f1} . Die erkennbare Symmetrie $\text{VAR}\{\alpha_{\rho}^*\} = \text{VAR}\{\alpha_{r-\rho}^*\}$ ist wegen der Erzeugung der Signale mit Hilfe einer

Schiebekette unmittelbar einleuchtend. Bei dem hier behandelten Signaltyp steigt die Varianz in allen Fällen zu den "mittleren" Koeffizienten hin an, wobei selbst die minimalen Werte in allen Fällen über denen ohne Berücksichtigung von Signal-Korrelationen liegen, wie die zum Vergleich in Bild 7.8a gestrichelt eingezeichneten Ergebnisse von Kap. 5.4.1 zeigen. Das ändert sich auch nicht bei dem Variablentyp x_{f3} (mit entgegengesetztem Vorzeichen der Kovarianzen $\psi_{x_{f3}x_{f3}}(1)$). Die Ergebnisse sind in Bild 7.8b aufgetragen.

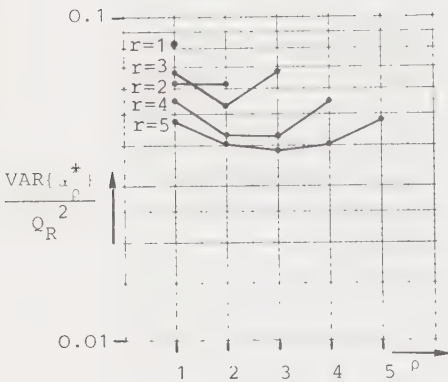


Zweierkomplement

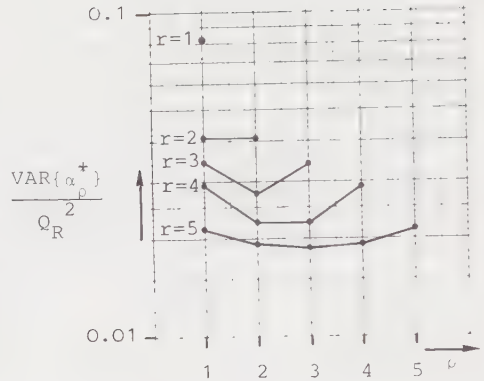


Modifizierte Form

Bild 7.8a: Varianzen der Rest-Koeffizientenfehler beim Signaltyp x_{f1} ($\rho_{\infty}=0.9$, $\psi_{\infty} = 0.3\pi + \psi_{x_{f1}x_{f1}}(1) > 0$)



Zweierkomplement



Modifizierte Form

Bild 7.8b: Varianzen der Rest-Koeffizientenfehler beim Signaltyp x_{f3} ($\rho_\infty=0.9$, $\psi_\infty=0.9\pi \rightarrow \psi_{x_{f3}x_{f3}}(1) < 0$)

Allerdings kehren sich hier die Verläufe über ρ um, d.h. infolge des Vorzeichenwechsels der Kovarianz benachbarter Variablen ($\psi_{x_{f3}x_{f3}}(1) < 0$, vgl. Bild 7.3c) steigen die Werte $\text{VAR}\{\alpha_\rho^*\}$ zu den "Rändern" hin an. Interessant ist der Fall $r = 3$: Die "Randwerte" übersteigen diejenigen von $r = 2$. Bei Variablen-Korrelationen kann somit nicht mehr allgemein angenommen werden, daß die Varianzen der Rest-Koeffizientenfehler monoton mit r fallen, wie es bei der Analyse für unkorrelierte Signale der Fall war (vgl. Bild 5.10b).

Schließlich fällt in Bild 7.8b gegenüber 7.8a auf, daß die Kurven für verschiedene r näher zusammenrücken: Das Maß der Reduktion der Werte $\text{VAR}\{\alpha_\rho^*\}$ hängt von den Korrelationen zwischen den Variablen ab; je stärker diese sind, desto geringer ist die Abnahme der Rest-Koeffizientenfehler-Varianzen mit steigendem r .

Abschließend sind einige Beispiele von Verteilungen der Rest-Koeffizientenfehler in den Bildern 7.9a-d zusammengestellt. Sie bestätigen im wesentlichen die Beobachtungen bei den Varianz-Untersuchungen: Breitere Verläufe der "inneren" Koeffi-

zientenfehler-Verteilungen bei $\varphi_{x_f x_f}^{(1)} > 0$, breitere Verläufe der "äußeren" bei $\varphi_{x_f x_f}^{(1)} < 0$. Darüber hinaus erkennt man bei einigen Fällen (Bild 7.9b,c,d) deutlich die Überschreitung des Wertebereiches $\left[-\frac{Q_R}{2}, \frac{Q_R}{2}\right]$, eine Erscheinung, die bereits bei der Untersuchung von Einzelbeispielen aufgefallen war.

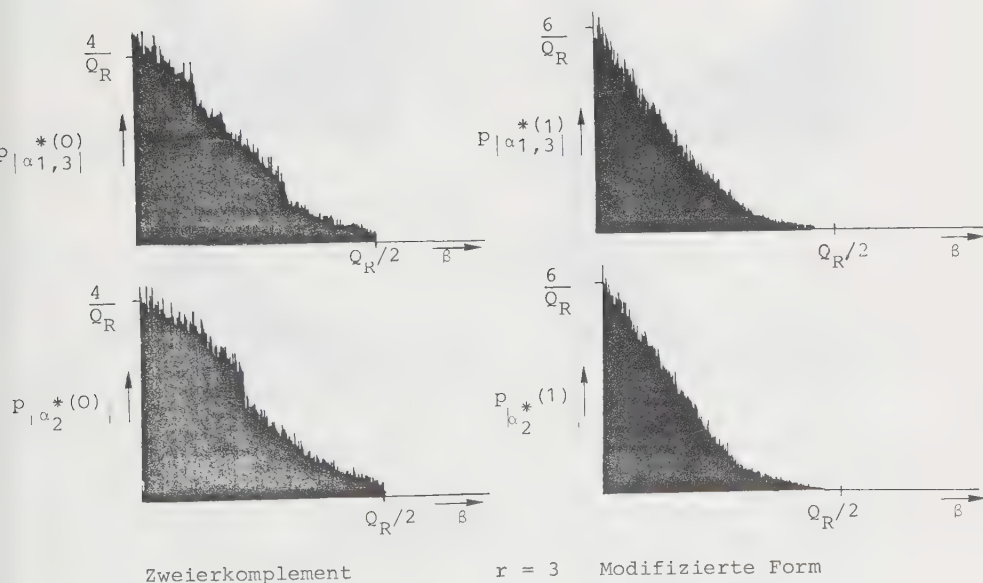


Bild 7.9a: Verteilung der Rest-Koeffizientenfehler
beim Signaltyp x_{f1} ($\rho_\infty = 0.9, \varphi_\infty = 0.3\pi$)

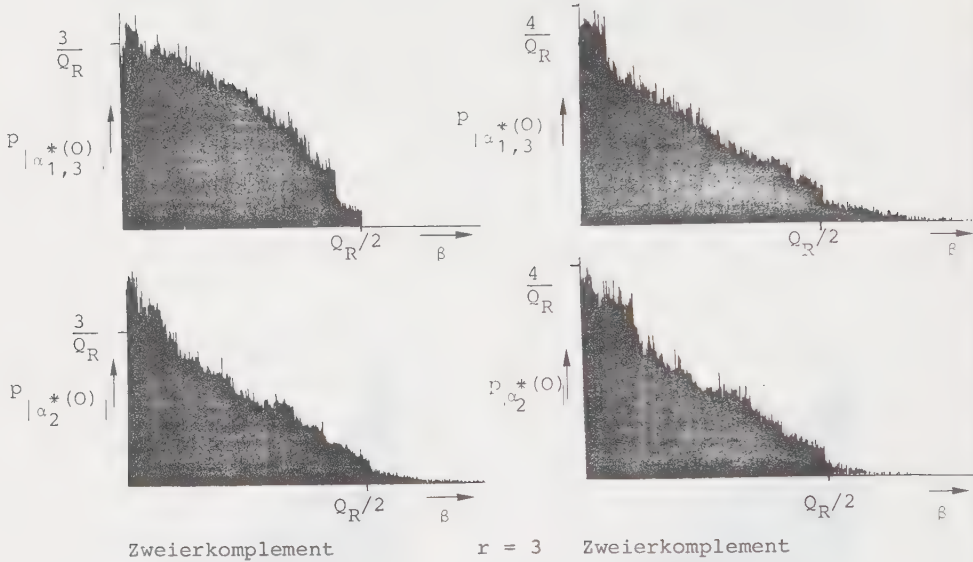
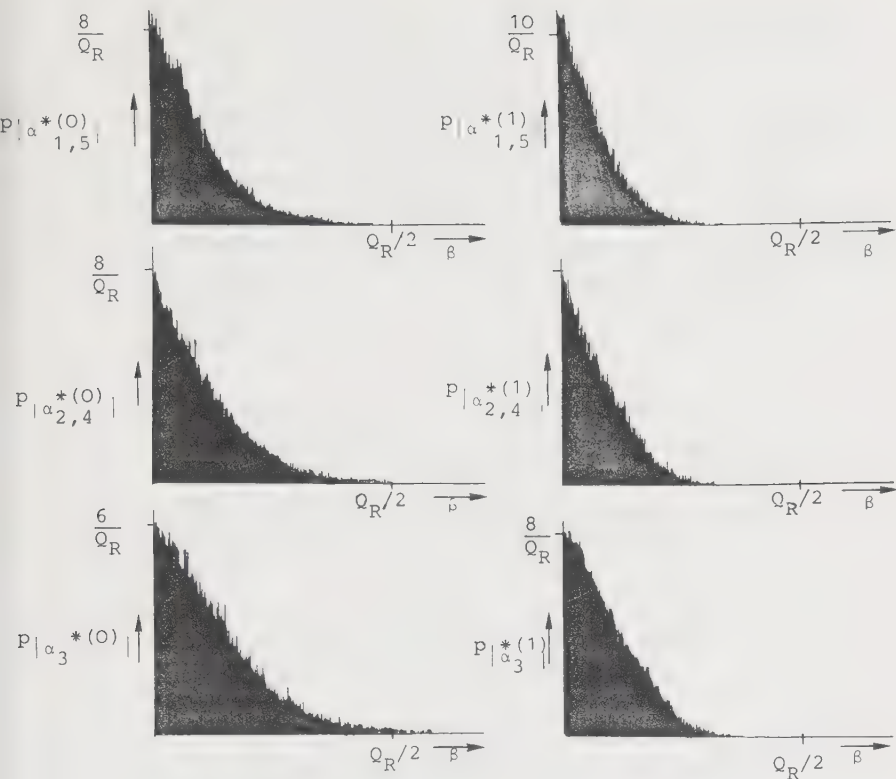


Bild 7.9b: Verteilungen der $\alpha_{\rho}^*(0)$ beim Signaltyp x_{f2}
 $(\rho_{\infty}=0.9, \psi_{\infty}=0.1\pi)$

Bild 7.9c: Verteilungen der $\alpha_{\rho}^*(0)$ beim Signaltyp x_{f3}
 $(\rho_{\infty}=0.9, \psi_{\infty}=0.9\pi)$



Zweierkomplement

$r = 5$

Modifizierte Form

Bild 7.9d: Verteilungen der Rest-Koeffizientenfehler
beim Signaltyp x_{f1} ($\rho_\infty = 0.9$, $\varphi_\infty = 0.3\pi$)

7.5 Abschließende Bemerkungen

Zusammenfassend läßt sich folgendes feststellen. Bei den Untersuchungen in diesem Kapitel wurden keinerlei Bedingungen an die statistischen Eigenschaften der Variablen gestellt. Eine geschlossene analytische Formulierung gelang dabei nicht mehr, jedoch ermöglichte die im Ansatz enthaltene Trennung von Signal- und Koeffizienten-Einfluß auf den Fehler die Untersuchung großer Ensembles und führte so zu statistischen Aussagen über das Fehlerverhalten. Die wichtigsten dabei gefundenen Ergebnisse seien noch einmal zusammengestellt:

1. Die Erwartungswerte der Gesamtfehlerleistungen sowie deren unkorrelierte Anteile werden gegenüber denen ohne Berücksichtigung von Variablen-Korrelationen reduziert. Das Maß der Reduktion hängt dabei von den spezifischen Variablen-Korrelationen ab.
2. Die Varianzen der Rest-Koeffizientenfehler nehmen infolge von Variablen-Korrelationen zu.
3. Die Rest-Koeffizientenfehler-Varianzen sind nun nicht mehr für alle r Koeffizienten gleich, sondern je nach den Vorzeichen der Kovarianzen benachbarter Signal-Abtastwerte können sie zur Mitte hin abfallen oder auch ansteigen.
4. Es trifft nicht mehr allgemein zu, daß mit steigendem r die Werte $\text{VAR}\{\alpha_{\rho}^*\}$ für alle ρ monoton fallen.
5. Der Wertebereich der Rest-Koeffizientenfehler kann den der a-priori-Koeffizientenfehler überschreiten.

Abschließend soll darauf hingewiesen werden, daß diese Aussagen auf der Basis von nur drei Beispielen von Signaltypen formuliert wurden. Deshalb muß eingeschränkt werden, daß Feststellungen wie die unter Punkt 1 und 2 nicht für jede beliebige Variablen-Statistik gelten müssen: Zur Überprüfung ihrer

Allgemeingültigkeit müßte eine weit größere Anzahl verschiedener Fälle untersucht werden.

8. ANWENDUNG DER FEHLERANALYSE AUF NICHTREKURSIVE FILTER

8.1 Allgemeine Vorbetrachtungen

Beim praktischen Aufbau nichtrekursiver Filter in direkter Form und Konzentrierter Arithmetik treten als hauptsächlichste Realisierungsungenauigkeiten lineare Verzerrungen infolge begrenzter Koeffizientenwortlängen auf, während rauschartige Fehler eine unbedeutende Rolle spielen [32,33,34].

In der vorliegenden Arbeit konnte gezeigt werden, daß bei Verwendung der Verteilten Arithmetik gerade die Koeffizientenfehler erheblich reduziert werden können, so daß diese Technik für eine Realisierung nichtrekursiver Filter besonders gut geeignet zu sein scheint. Allerdings erkauft man sich den Rückgang linearer Verzerrungen dabei mit einem Ansteigen von unkorrelierten, rauschartigen Fehlergrößen, so daß nun die Probleme der Koeffizienten-Quantisierungsfehler gegenüber der Konzentrierten Arithmetik verlagert sind.

Eine Schwierigkeit liegt nun darin, die beiden verschiedenartigen Fehlertypen - also deterministische und stochastische Fehler - miteinander zu vergleichen, etwa wenn Realisierungen in Konzentrierter und Verteilter Arithmetik einander gegenübergestellt und im Fehlerverhalten beurteilt werden sollen.

Eine Möglichkeit hierzu ist der Vergleich der gesamten Fehlerleistungen am Ausgang: Bei Erregung des Filters mit Rauschen äußern sich die Koeffizientenfehler ebenfalls als rauschartige Fehler im Ausgangssignal. Die Betrachtung der Gesamtfehlerleistung ist dann ein brauchbares Maß zur Beurteilung des Fehlerverhaltens, berücksichtigt allerdings nicht, inwieweit der Fehler mit dem Eingangssignal linear korreliert ist. Deshalb wird man eine andere Methode des Fehlervergleichs wählen müssen, wenn vornehmlich die linearen Verzerrungen, also z.B. die Verformungen des Frequenzganges interessieren.

Die Betrachtungen dieses letzten Kapitels unterteilen sich demgemäß in zwei Problemkreise: die Untersuchung der am Aus-

gang entstehenden Leistungen rauschartiger Fehler (Kap. 8.2) und die Bestimmung des realen Dämpfungsverlaufs des Filters bei begrenzter Koeffizientenspeicher-Wortlänge (Kap. 8.3). Zur Ausführung dieser Untersuchungen dient das Beispiel eines willkürlich gewählten nichtrekursiven Tiefpasses 47. Grades mit linearer Phase (Entwurf mit dem Programm von Parks/McClellan [35]). Bild 8.1a,b zeigt die zugehörige Impulsantwort und den Dämpfungsverlauf. Die folgenden Betrachtungen beziehen sich ausschließlich auf Erregung mit weißem, gleichverteiltem Rauschen.

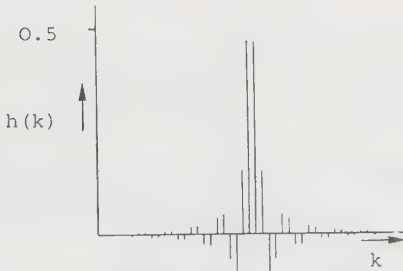


Bild 8.1a: Impulsantwort des betrachteten Filter-Beispiels

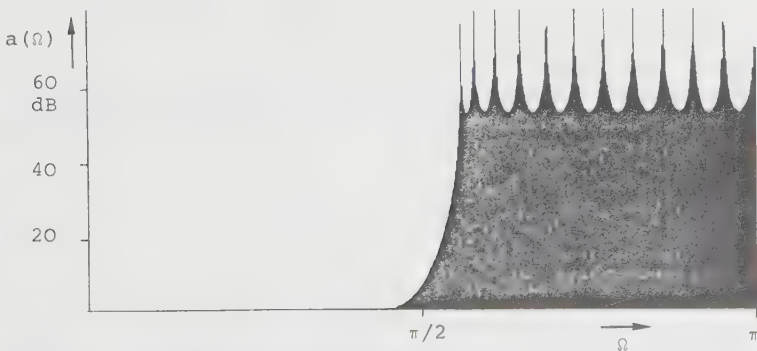


Bild 8.1b: Dämpfungsverlauf des betrachteten Filter-Beispiels
 (Durchlaßbereich: $a_{\max}=0.02\text{dB}$, $\Omega_D=0.44\pi$
 Sperrbereich: $a_{\min}=53\text{ dB}$, $\Omega_S=0.56\pi$)

8.2 Abhängigkeit des Signal-Störverhältnisses von der Quantisierungsstufe des Koeffizientenspeichers

Bei den allgemeinen Vorbetrachtungen zum prinzipiellen Fehlerverhalten verschiedener Vektor-Multiplizierer-Realisierungen (Kap. 4.1.1) wurden drei Fehlertypen unterschieden:

$\Delta y_E \hat{=}$ Fehler infolge Quantisierung der Eingangssignale

$\Delta y_A \hat{=}$ Fehler infolge Quantisierung des Ausgangssignals

und

Δy_c (Konzentrierte Arithmetik)	} $\hat{=}$	Fehler infolge endlicher Wortlänge des Koeffizientenspeichers
$\Delta y_{0,1}$ (Verteilte Arithmetik)		

In einer praktischen Realisierung sind im allgemeinen die Leistungen von Δy_E und Δy_A durch die A/D- und D/A-Wandler vorgegeben; d.h. diese Größen lassen sich nicht beliebig reduzieren, da wegen der prinzipiell beschränkten Genauigkeit analoger Schaltungselemente in den Wandlern Q_E und Q_A nicht beliebig klein gewählt werden können.

Die resultierende Ausgangs-Fehlerleistung besteht damit aus zwei durch Q_E und Q_A festliegenden Termen und einem weiteren Anteil, der durch die Wahl der Quantisierungsstufe im Inneren des Filters - im Fall der Verteilten Arithmetik also durch die Quantisierungsstufe des Koeffizientenspeichers Q_R - beeinflusbar ist:

$$N_{\text{res}} = \frac{Q_E^2}{12} R_E + \frac{Q_A^2}{12} + N_{\Delta y}(Q_R) \quad (8.1a)$$

mit der Rauschzahl

$$R_E = \sum_{\kappa=0}^{R-1} h^2(\kappa) = \sum_{\rho=1}^R a_{\rho}^2 = \frac{1}{\pi} \int_0^{\pi} |H(\Omega)|^2 d\Omega. \quad (8.1b)$$

Daraus wird deutlich, daß die Reduktion von Q_R nur bis zu einer gewissen Grenze hin sinnvoll ist, denn die resultie-

rende Ausgangsfehlerleistung kann den Wert $\frac{Q_E^2}{12} R_E + \frac{Q_A^2}{12}$ grundsätzlich nicht unterschreiten.

Zur quantitativen Beschreibung führt man dazu unter Annahme von $Q_E = Q_A$ eine innere Wortlängenerhöhung

$$\Delta w = 1d \frac{Q_E}{Q_R} \quad (\Delta w \in N_0) \quad (8.2)$$

ein, die bei Erfüllung der Bedingung

$$N_{\Delta Y}(Q_R) \leq \frac{Q_E^2}{12} (R_E + 1) \quad (8.3)$$

zu einem sinnvollen Kompromiß zwischen Realisierungsaufwand und Fehlerleistungs-Reduktion führt [36]. Diese Werte Δw sollen im folgenden für das oben eingeführte Beispiel eines nichtrekursiven Tiefpasses vom Grad $R-1 = 47$ für verschiedene Koeffizientengruppierungen ($r = 2, 4, 8$) bestimmt werden. Dazu berechnet man die auf die gesamte Ausgangssignalleistung

$$N_A = \sigma_x^2 \sum_{\kappa=0}^{R-1} h^2(\kappa) = \frac{1}{3} \cdot R_E$$

bezogene resultierende Fehlerleistung, also den Stör-Signalabstand:

$$N' = N_{res}/N_A = \frac{Q_E^2}{4} \left(1 + \frac{1}{R_E}\right) + 3 N_{\Delta Y}/R_E \quad (8.4)$$

Im gewählten Beispiel ist $R_E = 0.48522$, die äußeren Quantisierungsstufen werden mit $Q_E = Q_A = 2^{-7}$ festgelegt.

Für die Größen $N_{\Delta Y}$ in (8.4) werden nun zunächst nicht die individuellen Größen für den hier verwendeten speziellen Koeffizientensatz benutzt, die sich für jeden Wert Q_R nach Kap. 4.4 aus den a-priori-Koeffizientenfehlern und den Fehlervektoren ergeben, sondern man setzt statt dessen die in Kapitel 5.3 gefundenen Erwartungswerte ein, die auf der Basis großer Ensembles mit gleichverteilten a-priori-Koeffizientenfehlern bestimmt wurden:

$$N_{\Delta Y} \leftarrow \frac{R}{r} \cdot E\{N_{\Delta Y}\} \Big|_r \quad (8.5)$$

Damit erhält man die in den Bildern 8.1a,...,d dargestellten Verläufe der resultierenden Fehlerleistung bzw. deren unkorrelierten Anteil als Funktion der Koeffizientenspeicher-Quantisierungsstufe.

Die Kurven fallen für kleine innere Wortlängen zunächst mit 6 dB/Bit ab entsprechend dem hier überwiegenden, zu Q_R^2 proportionalen Anteil $\frac{R}{r} E\{N_{\Delta Y}\} \Big|_r$ und laufen dann schließlich auf einen konstanten Wert ein, der nur noch durch die externen Quantisierungsstufen Q_E und Q_A bestimmt wird:

$$N' \rightarrow \frac{2^{-14}}{4} \left(1 + \frac{1}{0.48522}\right) = -43.306 \text{ dB.}$$

Die Werte für Δw sind diesen Kurven zu entnehmen. Sie betragen beim Zweierkomplement für $r = 4$ und $r = 8 \rightarrow \Delta w = 2$ Bit, für $r = 2 \rightarrow \Delta w = 1$ Bit falls nur der unkorrelierte Rauschanteil betrachtet wird oder ebenfalls $\Delta w = 2$ Bit für die Gesamtleistung. Die modifizierte Form liegt wie zu erwarten günstiger: Mit Ausnahme von $r = 2$ genügt in allen Fällen eine Vergrößerung der Wortlänge im Inneren um 1 Bit. Es muß darauf hingewiesen werden, daß die willkürliche Wahl der äußeren Quantisierungsstufen $Q_E = Q_A = 2^{-7}$ keine Einschränkung der Allgemeinheit bedeutet: Die hier gefundenen Werte Δw ergeben bei beliebig vorgegebenen Werten $Q_E = Q_A$ die technisch sinnvollen Quantisierungsstufen für den Koeffizientenspeicher.

Für die Werte $N_{\Delta Y}$ wurden zunächst die Erwartungswerte aus den Ensemble-Untersuchungen gesetzt, in der Hoffnung, die verschiedenen Koeffizientengruppen würden bereits zu einem Mittelungseffekt führen, durch den die Erwartungswerte hinreichend gut getroffen werden. Die folgenden Untersuchungen sollen klären, inwieweit dies der Fall ist: Die Verläufe der Erwartungswerte werden denen nach den individuellen Analysen bei den jeweiligen Quantisierungsstufen Q_R gegenübergestellt (Bild 8.2a...f, Bild 8.3a...f).

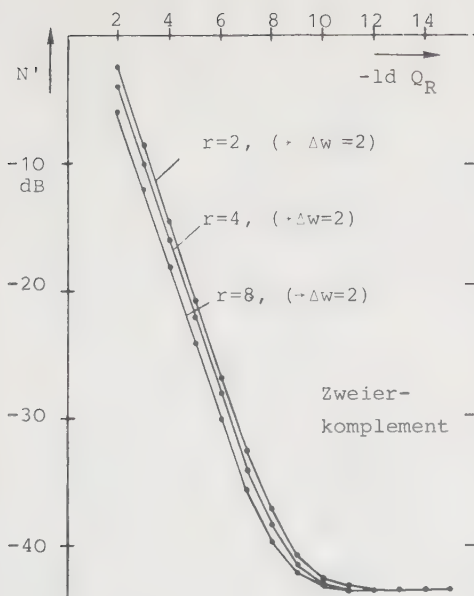


Bild 8.1a:

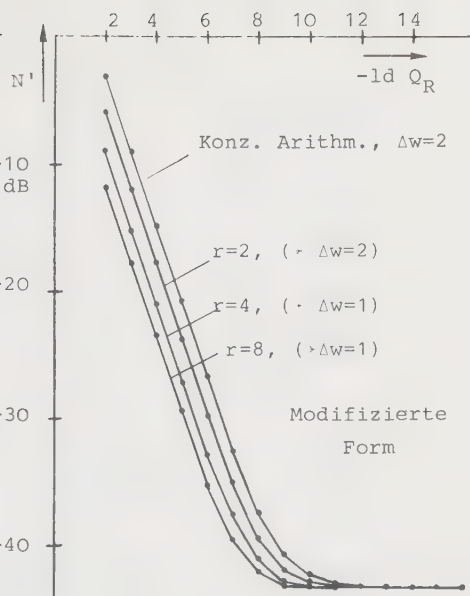


Bild 8.1b:

Stör-Signalverhältnis bei Betrachtung der gesamten Fehlerleistung

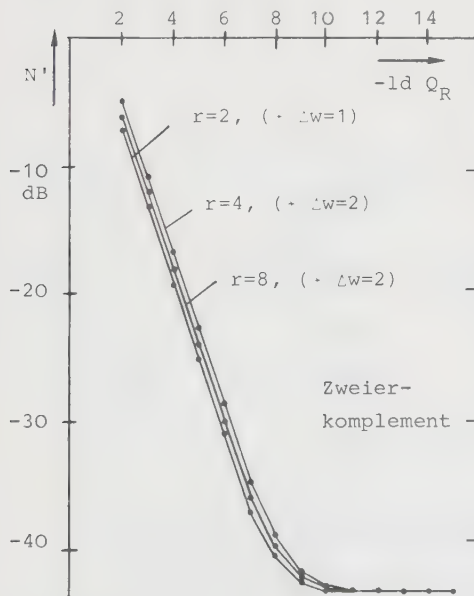


Bild 8.1c:

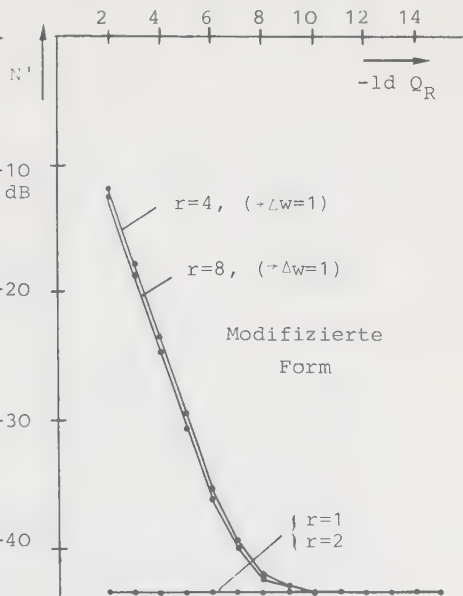


Bild 8.1d:

Stör-Signalverhältnis bei Betrachtung des unkorrelierten Fehleranteils

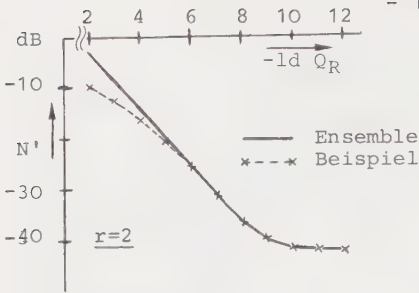


Bild 8.2a

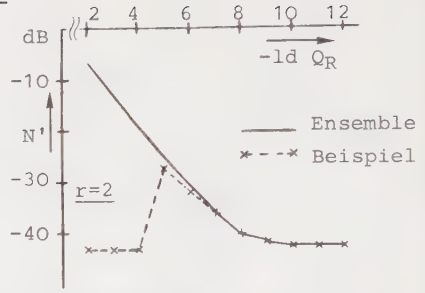


Bild 8.2d

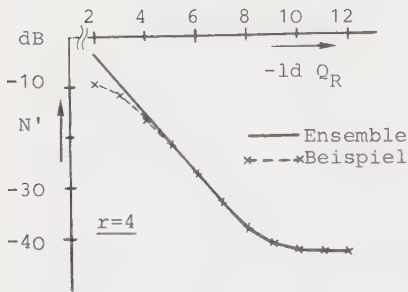


Bild 8.2b

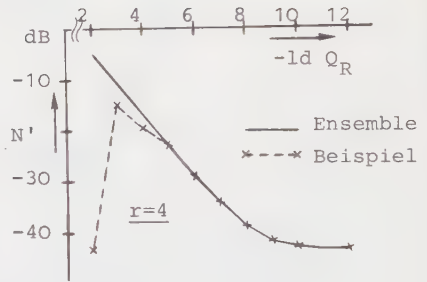


Bild 8.2e

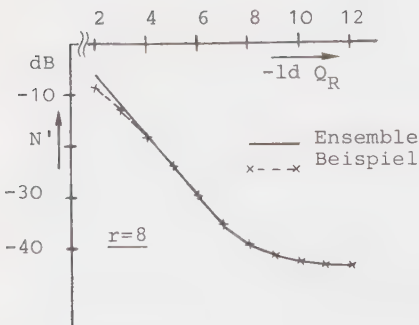


Bild 8.2c

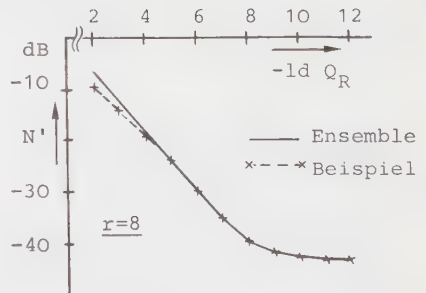


Bild 8.2f

Bild 8.2a...f: Stör-Signalabstand beim Zweierkomplement

(Bild 8.2a...c: Betrachtung der Gesamt-Fehlerleistung

Bild 8.2d...f: Betrachtung des unkorrelierten Rauschanteils)

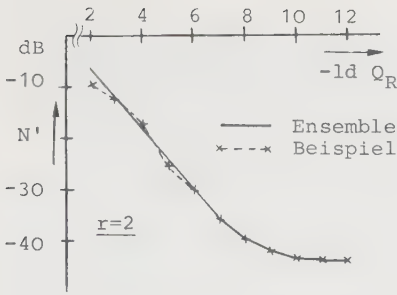


Bild 8.3a

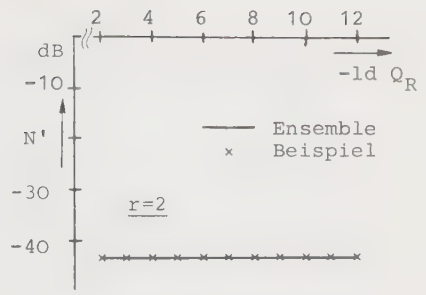


Bild 8.3d

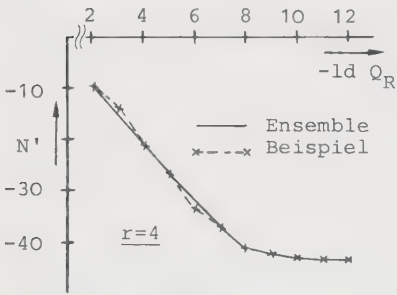


Bild 8.3b

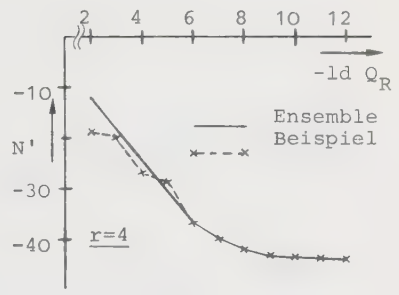


Bild 8.3e

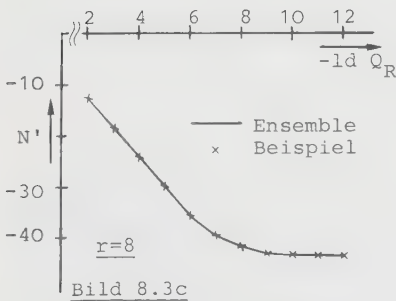


Bild 8.3c

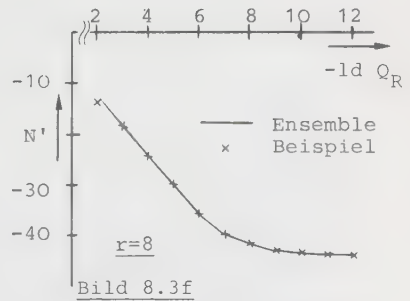


Bild 8.3f

Bild 8.3a...f: Stör-Signalabstand bei modifizierter Form

(Bild 8.3a...c: Betrachtung der Gesamt-Fehlerleistung

Bild 8.3d...f: Betrachtung des unkorrelierten Rauschanteils

Es zeigt sich bei allen betrachteten Koeffizienten-Gruppierungen für Werte $Q_R \leq 2^{-5}$ eine sehr gute Übereinstimmung. Die geringe Anzahl von $\frac{1}{2} R/r$ Koeffizientengruppen *) reicht hier offenbar aus, um zur Berechnung der Ausgangs-Fehlerleistung bei nichtrekursiven Filtern die individuellen Werte $N_{\Delta y}$ durch deren Erwartungswerte zu ersetzen (vgl. Kap. 5.3, Bild 5.5). Dabei wirkt sich die folgende Tatsache günstig aus: Bei großen Werten r steht zur Mittelung eine geringere Anzahl von Gruppen zur Verfügung als bei kleinem r (z.B. $r = 8 \rightarrow 3$ Gruppen; $r = 2 \rightarrow 12$ Gruppen). Die Ensemble-Untersuchungen in Kapitel 5.3 haben jedoch gezeigt, daß die Streuungen der Fehlerleistungen mit steigendem r fallen, so daß also bei großen Werten r ein geringerer Ensemble-Umfang erforderlich ist als bei kleinen, um mit gleicher Wahrscheinlichkeit in die Nähe des Erwartungswertes zu kommen.

Für Werte $Q_R > 2^{-5}$, die praktisch nicht von Bedeutung sind, ergeben sich deutliche Unterschiede zwischen den mit den individuellen Werten $N_{\Delta y}$ und deren Erwartungswerten berechneten Fehlerleistungen. Die Erklärung liegt darin, daß sich hier die Quantisierungsstufen Q_R dem Aussteuerungsbereich der Koeffizienten (vor allem an den "Rändern"!) annähern. Damit trifft die Annahme von gleichverteilten, von den Koeffizienten statistisch unabhängigen a-priori-Koeffizientenfehlern nicht mehr zu.

An einem Beispiel soll das näher erläutert werden. Man betrachte für den Fall $r = 4$ eine der 12 Koeffizientengruppen des hier behandelten Filter-Beispiels (Es sei vermerkt, daß die Zusammenfassung von Impulsantwort-Abtastwerten unter Einhaltung maximaler zeitlicher Abstände, also gemäß der Multiplex-Realisierung nach Kap. 3.2 erfolgte):

*) Wegen Linearphasigkeit sind jeweils 2 Gruppen paarweise gleich.

$$a_5 = h(1) = -0.00183$$

$$a_6 = h(13) = -0.01720$$

$$a_7 = h(25) = 0.14903$$

$$a_8 = h(37) = 0.00890$$

Für die Quantisierungsstufe $Q_R = 2^{-2}$ ergeben sich die a-priori Koeffizientenfehler

$$\alpha_5 = -0.00183 = -0.00732 Q_R$$

$$\alpha_6 = -0.01720 = -0.06880 Q_R$$

$$\alpha_7 = -0.10097 = -0.40388 Q_R$$

$$\alpha_8 = 0.00890 = 0.03560 Q_R ,$$

von denen α_5, α_6 und α_8 mit den zugehörigen Koeffizienten selbst übereinstimmen; d.h. $\begin{bmatrix} a_5 \\ a_6 \\ a_8 \end{bmatrix} Q_R = \begin{bmatrix} a_5 \\ a_6 \\ a_8 \end{bmatrix} Q_R = \begin{bmatrix} a_5 \\ a_6 \\ a_8 \end{bmatrix} Q_R = 0$

Setzt man die Werte in Gl. (4.28a) ein, so ergibt sich beim Zweierkomplement für die Fehlervektor-Komponenten

$$e_n^{(0)} = 0 \quad \forall n \in \{0, \dots, 2^F - 1\},$$

d.h. mit Gl. (4.68a) verschwindet der unkorrelierte Fehleranteil N_{ϵ_0} .

Da sich die gleichen Verhältnisse für alle übrigen Koeffizientengruppen ergeben, besteht der unkorrelierte Anteil des resultierenden Ausgangsfehlers nur aus den durch Eingangs- und Ausgangsquantisierung hervorgerufenen Rauschfehlern; Bild 8.2e macht dieses deutlich. Durch Einsetzen des Erwartungswertes

$$N_{\epsilon} \leftarrow E\{N_{\epsilon}\} = 0.0715 Q_R^2$$

macht man also in diesem Falle einen erheblichen Fehler.

Es wurde bereits darauf hingewiesen, daß die geringen Koeffizientenspeicher-Wortlängen, bei denen derartige Diskrepanzen

auftreten, im allgemeinen technisch nicht interessant sind; d.h. für praktische Fälle kann in der Regel mit den Erwartungswerten gerechnet werden. Diese Tatsache ist deshalb von Bedeutung, weil sich damit allgemeine Aussagen über die erforderlichen Wortlängen-Erhöhungen des Koeffizientenspeichers bei nichtrekursiven Filtern formulieren lassen.

Dazu wird in Gl. (8.3) die Größe $N_{\Delta Y}$ durch ihren Erwartungswert ersetzt, der zu Q_R^2 proportional ist:

Mit der auf Q_R^2 normierten Fehlerleistung

$$E\{n_{\Delta Y}\} = \frac{1}{Q_R^2} \frac{R}{r} \cdot E\{N_{\Delta Y}\} \Big|_r \quad (8.6a)$$

und der Bedingung

$$E\{n_{\Delta Y}\} \cdot Q_R^2 \leq \frac{Q_E^2}{12} (R_E + 1) \quad (8.6b)$$

erhält man aus (8.2) für die zu fordernde Wortlängen-Erhöhung

$$\Delta w = \lg \frac{Q_E}{Q_R} \geq \frac{1}{2} \lg \frac{12 E\{n_{\Delta Y}\}}{(R_E + 1)}, \quad \Delta w \in \mathbb{N}_0. \quad (8.7)$$

Sie hängt bei vorgegebener Koeffizienten-Gruppierung ($\frac{R}{r}$ Gruppen der Länge r) offenbar nur noch von der Rauschzahl R_E ab. Wählt man nun, wie in vielen Fällen üblich, die Filter-Koeffizienten so, daß der Maximalwert des Frequenzgangs "1" wird, dann liegt R_E gemäß Gl. (8.1b) je nach Filter-Charakteristik zwischen "0" und "1". Aus (8.7) ergibt sich damit für $R_E = 0$ die maximal, für $R_E = 1$ die mindestens erforderliche Erhöhung der inneren Wortlänge. Demgemäß sind in den Bildern 8.4a...f in Abhängigkeit vom Filtergrad die für Erregung mit weißem, gleichverteiltem Rauschen **a l l g e m e i n g ü l t i g e n o b e r e n u n d u n t e r e n G r e n z e n d e r W e r t e Δw** dargestellt, und zwar für gewöhnliches Zweierkomplement (Bild 8.4a,b,c) und modifizierte Form (Bild 8.4d, e,f) bei verschiedenen Koeffizientengruppen-Längen r . Die jeweiligen Werte für

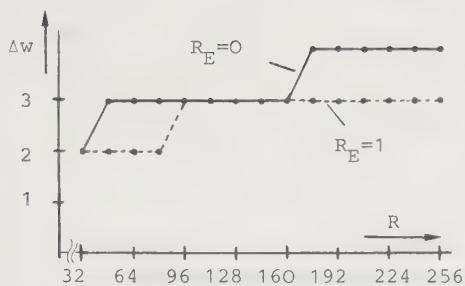


Bild 8.4a: Zweierkomplement $r=2$

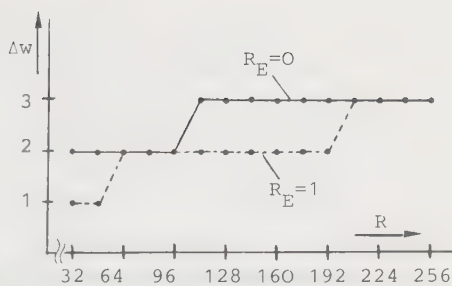


Bild 8.4d: Modifizierte Form $r=2$

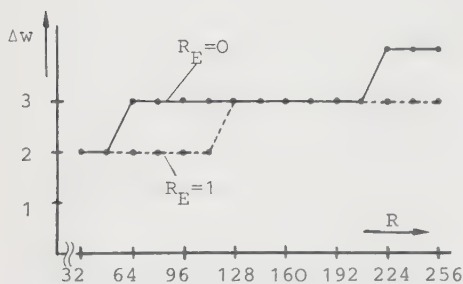


Bild 8.4b: Zweierkomplement $r=4$

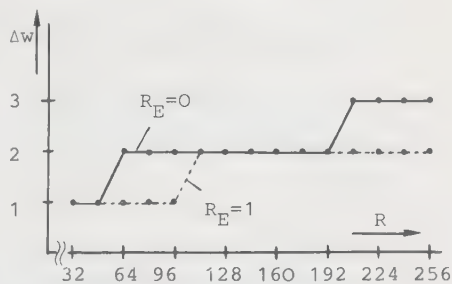


Bild 8.4e: Modifizierte Form $r=4$

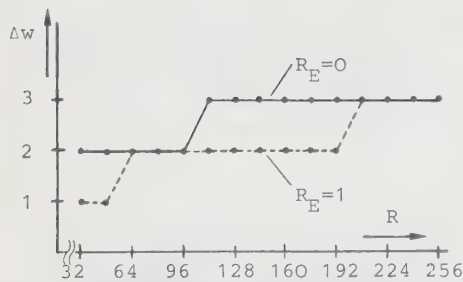


Bild 8.4c: Zweierkomplement $r=8$

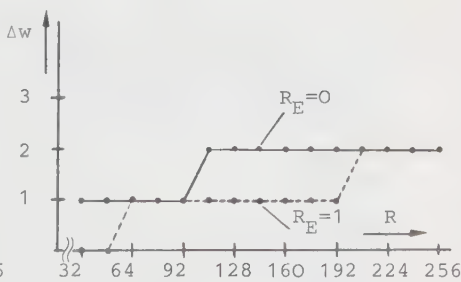


Bild 8.4f: Modifizierte Form $r=8$

Bild 8.4a...f: Maximal und minimal erforderliche Erhöhung der Wortlänge im Inneren nichtrekursiver Filter als Funktion der Filterlänge R . (Erregung der Filter mit weißem, gleichverteiltem Rauschen).

$$E\{n_{\Delta Y}\} = \frac{R}{r} \left(\frac{1}{Q_R} E\{N_{\Delta Y}\} \mid r \right)$$

werden dabei dem Bild 5.5 entnommen.

8.3 Verzerrungen des Dämpfungsverlaufs nichtrekursiver Filter infolge endlicher Koeffizientenspeicherwortlänge

8.3.1 Definition des Frequenzgangs

Der Frequenzgang nichtrekursiver Filter berechnet sich direkt aus der Impulsantwort nach Gl. (8.8)

$$H(\Omega) = \sum_{\rho=0}^{R-1} h(\rho) e^{-j\rho\Omega} \quad (8.8)$$

In einer praktischen Realisierung in Konzentrierter Arithmetik sind hier die idealen Werte $h(\rho)$ durch die auf die Quantisierungsstufe Q_C des Koeffizientenspeichers gerundeten Werte $[h(\rho)]_{Q_C}$ zu ersetzen, was eine Verzerrung des Frequenzgangs gegenüber der idealen Ausgangslösung bewirkt. Bild 8.5 zeigt den realen Dämpfungsverlauf des in Kap. 8.1 eingeführten Tiefpaß-Beispiels bei einer Quantisierungsstufe $Q_C = 2^{-7}$ (vgl. den idealen Dämpfungsverlauf nach Bild 8.1b).

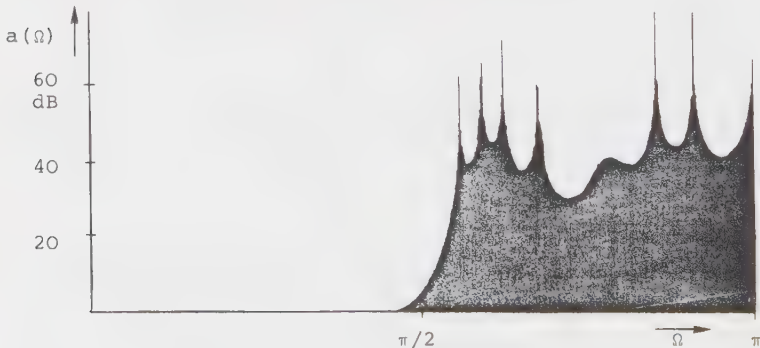


Bild 8.5: Dämpfungsverlauf nach Rundung der Koeffizienten auf $Q_C = 2^{-7}$.

Für die Verteilte Arithmetik ist eine Definition des Frequenzgangs nach Gl. (8.8) ungeeignet, denn nach den Untersuchungen der vorangegangenen Kapitel führt hier eine Wortlängenbegrenzung des Koeffizientenspeichers zu nichtlinearem Charakter des Systems, was sich als (nun reduzierte) lineare Verzerrung und ein zusätzlicher rauschartiger Fehler äußert. Interessiert man sich dennoch für den realen Dämpfungsverlauf des in Verteilter Arithmetik realisierten nichtrekursiven Filters, so ist es zweckmäßig, diesen entsprechend einer praktischen Messung im Zeitbereich zu formulieren.

Eine Möglichkeit hierzu ist die Erregung des Filters mit einer Sinusfolge der normierten Frequenz Ω , die schrittweise von 0 bis π variiert wird. Das Ausgangssignal wird dabei selektiv bei der jeweiligen Eingangsfrequenz gemessen. Gegen dieses Verfahren spricht, daß die Erregung mit einzelnen Sinusfunktionen im allgemeinen keinem realen Betriebsfall entspricht. Da aber, wie in den vorangegangenen Abschnitten gezeigt wurde, das Fehlerverhalten wesentlich von den Eigenschaften der Eingangssignale abhängt, verliert eine derartige Frequenzgang-Bestimmung bei der Verarbeitung realistischer Signal-Gemische seine Gültigkeit.

Es erscheint daher zweckmäßig, von einer Erregung mit weißem Rauschen auszugehen. Ein solches Vorgehen kann zwar auch nicht repräsentativ für alle möglichen Filterungsprobleme gesetzt werden, stellt aber zumindest einen denkbaren praktischen Fall dar.

Der Dämpfungsgang ist auf diese Weise leicht anzugeben, indem das Leistungsspektrum am Ausgang bestimmt wird; dieses beschreibt dann die wirksame Dämpfung des Systems bei Erregung mit spektral gleichverteiltem Rauschen.

Auf der Basis dieser Definition des Filter-Dämpfungsgangs sollen die folgenden Untersuchungen eines nichtrekursiven Filters in Verteilter Arithmetik durchgeführt werden. Es sei darauf hingewiesen, daß Phasenverzerrungen des Filters hier nicht betrachtet werden.

8.3.2 Berechnung der Dämpfung durch Simulation

Zur Bestimmung des Ausgangs-Leistungsspektrums wird folgendes Blockdiagramm programmiert:

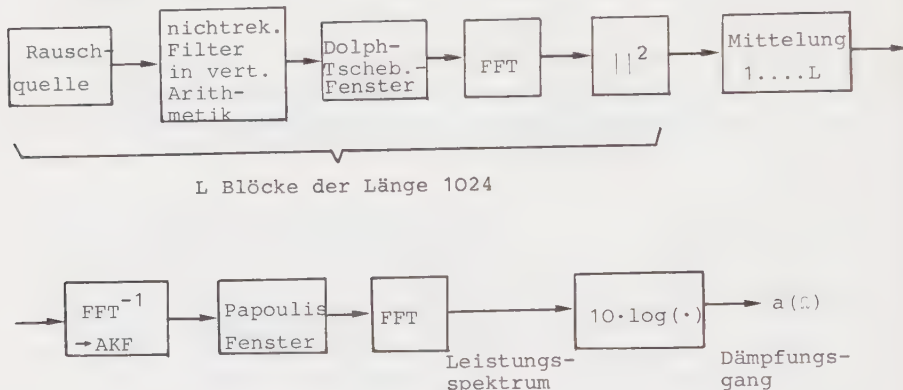


Bild 8.6: Blockdiagramm zur Berechnung des realen Dämpfungs-gang durch Simulation

Eine Rauschquelle liefert dabei Abtastwerte von gleichverteiltem weißem Rauschen, und zwar insgesamt L Blöcke mit jeweils einer Länge von 1024 Werten. Die einzelnen Leistungsspektren der L Filterausgangs-Signalabschnitte werden gemittelt. Um die Spektralverzerrungen infolge der Rechteck-Fensterung der einzelnen Signalabschnitte zu reduzieren, wird jeder Block vor der FFT einer Dolph-Tschebyscheff-Fensterung unterworfen [37]. Der so erhaltene Verlauf des Leistungsspektrums wird anschließend geglättet, indem die durch Rücktransformation berechnete Autokorrelationsfunktion mit einem Papoulis-Fenster multipliziert wird [38].

Auf diese Weise erhält man die in den Bildern 8.7a,...,e (Zweierkomplement) und 8.8a,...,e (modifizierte Form) wiedergegebenen realen Filter-Dämpfungsverläufe. In den Simulationsläufen wurden dabei jeweils $L = 20$ Blöcke überlagert. Die Quantisierungsstufe der Koeffizientenspeicher betrug $Q_R = 2^{-7}$. Eine Wortlängenbegrenzung des Ausgangssignals wurde nicht durchgeführt, um nur die Auswirkungen der Koeffizienten-

speicher-Quantisierung zu erfassen.

Die Dämpfungsverzerrungen sind deutlich im Sperrbereich zu beobachten: Für kleine Werte r ergibt sich noch eine starke Welligkeit, die von den relativ großen linearen Verzerrungen, also den Rest-Koeffizientenfehlern herrühren. Für größere r ebnet sich der Dämpfungsgang immer mehr ein, d.h. es verbleibt schließlich bei $r = 8$ ein rauschartiger, spektral gleichverteilter Fehler. Gemäß den Untersuchungen in Kap. 4 und 5 beträgt er bei gewöhnlicher Zweierkomplement-Codierung $\frac{R}{r} \cdot \frac{Q_R^2}{9}$ (vgl. Gl. 5.11a bzw. Bild 5.7) und bewirkt damit im Sperrbereich eine Dämpfung

$$a(\Omega) = -10 \log \left(\frac{R}{r} \frac{Q_R^2}{9} / \sigma_x^2 \right) = -10 \log \left(\frac{2}{3} 2^{-14} / \frac{1}{3} \right) = 39.1 \text{ dB}$$

$$0.56\pi \leq \Omega \leq \pi$$

$$(r = 8, Q_R = 2^{-7}, \sigma_x^2 = \frac{1}{3}).$$

Für die modifizierte Form ergibt sich entsprechend

$$a(\Omega) = -10 \log \left(\frac{R}{r} \frac{Q_R^2}{36} / \sigma_x^2 \right) = 45.1 \text{ dB};$$

$$0.56\pi \leq \Omega \leq \pi$$

Allgemein kann festgestellt werden, daß in beiden Fällen durch die spektral gleichmäßigere Verteilung bei großen r die wirksame Sperrdämpfung gegenüber der bei Konzentrierter Arithmetik erheblich vergrößert wird. Bei der modifizierten Form ist gegenüber dem Zweierkomplement ein Gewinn von 6 dB zu verzeichnen (vgl. die Bilder 8.7, 8.8 und 8.5).



Bild 8.7a

$r = 2$

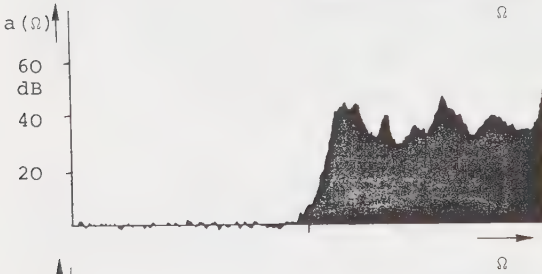


Bild 8.7b

$r = 3$



Bild 8.7c

$r = 4$

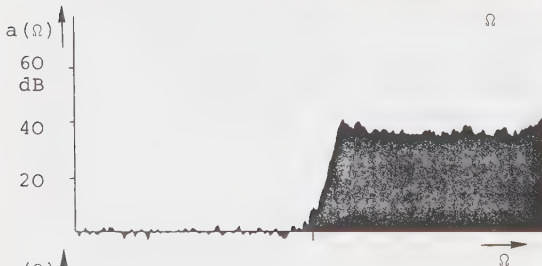


Bild 8.7d

$r = 6$



Bild 8.7e

$r = 8$

Dämpfungsverläufe bei gewöhnlichem Zweierkomplement ($Q_R = 2^{-7}$)

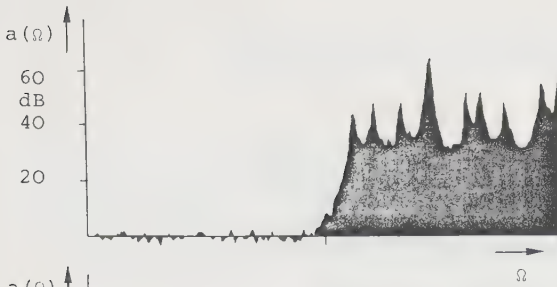


Bild 8.8a
 $r = 2$

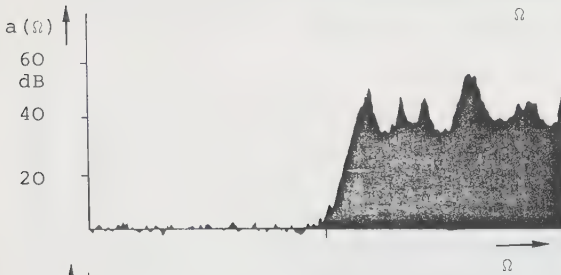


Bild 8.8b
 $r = 3$

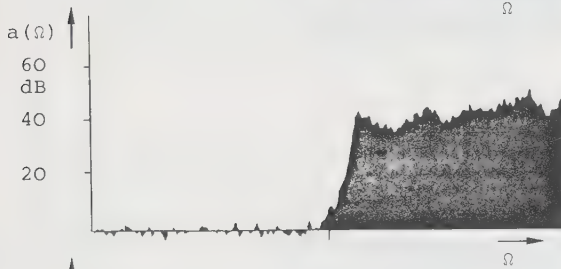


Bild 8.8c
 $r = 4$



Bild 8.8d
 $r = 6$

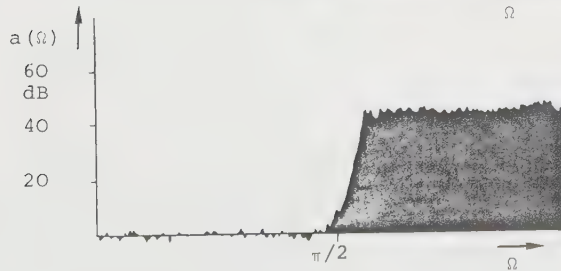


Bild 8.8e
 $r = 8$

Dämpfungsverläufe bei der modifizierten Form ($Q_R = 2^{-7}$)

8.3.3 Analytische Behandlung mit dem linearen Ersatzmodell

Die Berechnung der Rest-Koeffizientenfehler ist nach Kap. 4.6 für weißes, gleichverteiltes Eingangsrauschen geschlossen möglich (Gl. 4.66). Nach den dort zugrunde liegenden Vorstellungen entsteht neben den linearen Verzerrungen infolge α_{ρ}^* ein unkorrelierter rauschartiger Fehleranteil, der demnach als spektral weiß anzunehmen ist.

Es liegt also nahe, die im letzten Abschnitt durch Simulation ermittelten Dämpfungsverläufe analytisch zu konstruieren, indem zunächst die Koeffizienten der idealen Ausgangslösung um die Rest-Koeffizientenfehler "verfälscht" werden, woraus ein verzerrter Frequenzgang zu berechnen ist. Seinem quadrierten Betrag ist dann die Leistung des unkorrelierten Fehlers

$$N_{\epsilon} = \frac{1}{\sigma_x^2} \sum_{\ell=1}^{R/r} N_{\epsilon}^{(\ell)}$$

(Gl. 4.68) additiv hinzuzufügen, um den Vergleich mit den Simulationsergebnissen durchführen zu können; man erhält das Ausgangs-Leistungsspektrum und daraus den Dämpfungsverlauf:

$$a(\Omega) = -10 \log \left(\left| \sum_{\rho=1}^R (a_{\rho} - \alpha_{\rho}^*) e^{-j(\rho-1)\Omega} \right|^2 \right) + \frac{1}{\sigma_x^2} \sum_{\ell=1}^{R/r} N_{\epsilon}^{(\ell)} \quad (8.9)$$

mit $a_{\rho} = h(\rho-1)$.

In den Bildern 8.9a,b,c und 8.10a,b,c sind für einige Beispiele die nach Gl. (8.9) gewonnenen Ergebnisse wiedergegeben.

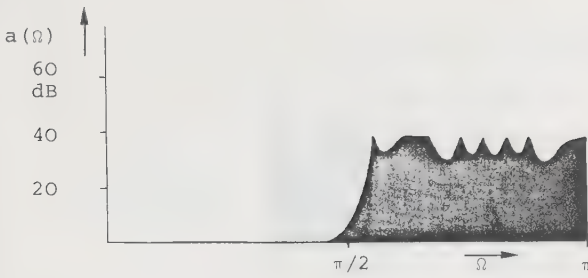


Bild 8.9a

$r = 2$

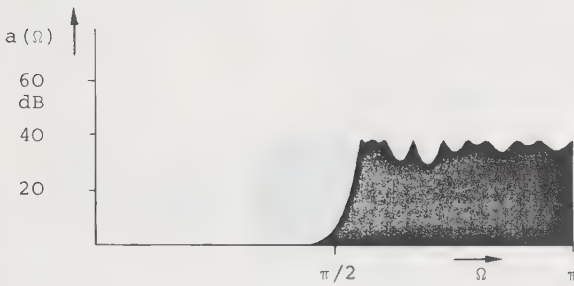


Bild 8.9b

$r = 3$

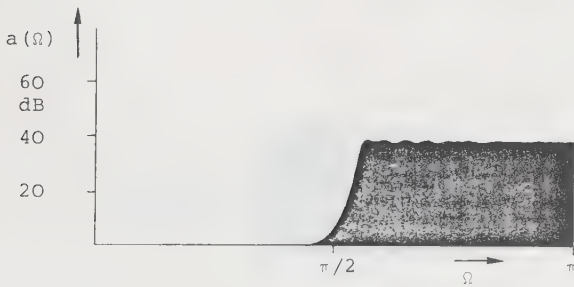


Bild 8.9c

$r = 8$

Bild 8.9: Dämpfungsverläufe nach linearem Ersatzmodell
(Zweierkomplement)

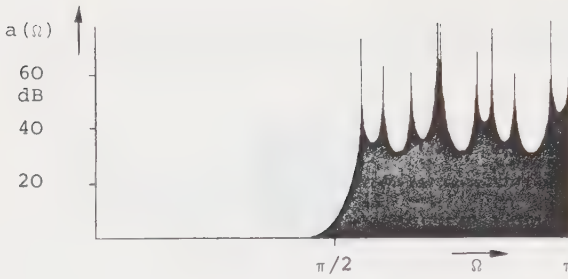


Bild 8.10a
 $r = 2$

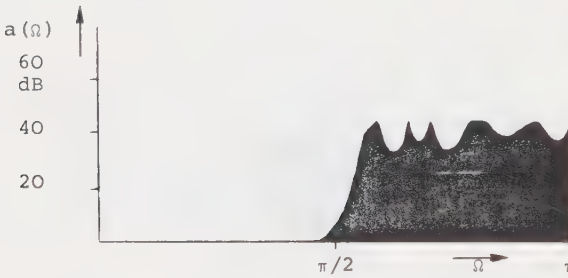


Bild 8.10b
 $r = 3$

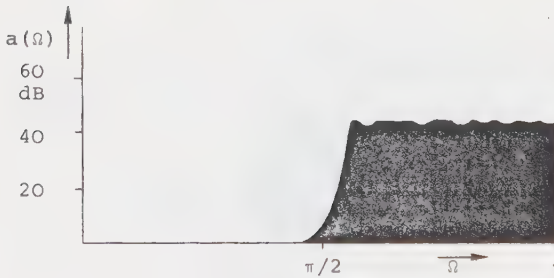


Bild 8.10c
 $r = 8$

Bild 8.10: Dämpfungsverläufe nach linearem Ersatzmodell
(modifizierte Form)

8.3.4 Vergleich von Analyse und Simulation

Zum Vergleich sind in den Bildern 8.11a,b,c und 8.12a,b,c die Dämpfungsverläufe nach Simulation und analytischer Berechnung übereinandergezeichnet, und zwar für die Beispiele $r = 2, 3$ und 8 bei gewöhnlichem und modifiziertem Zweierkomplement.

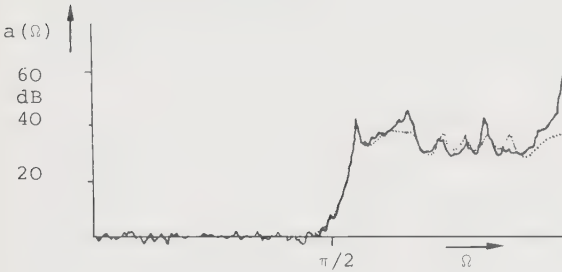


Bild 8.11a

$r = 2$

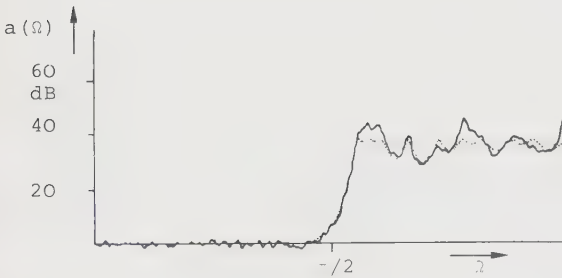


Bild 8.11b

$r = 3$

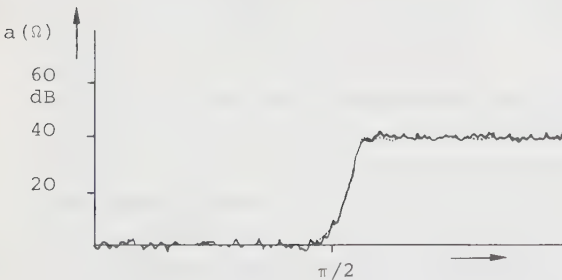


Bild 8.11c

$r = 8$

Bild 8.11: Vergleich der Dämpfungsverläufe beim Zweierkomplement

———— Simulation

..... Berechnung nach linearem Ersatzmodell

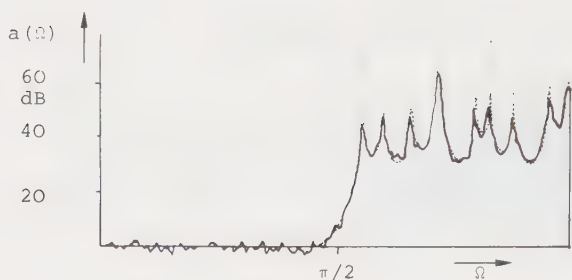


Bild 8.12a

$r = 2$

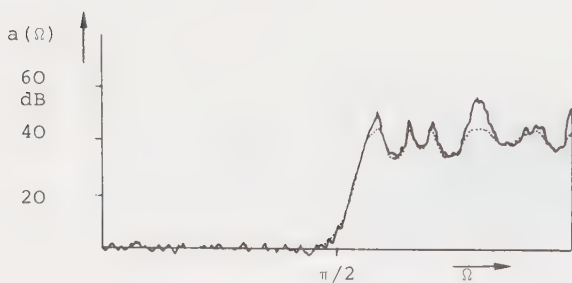


Bild 8.12b

$r = 3$

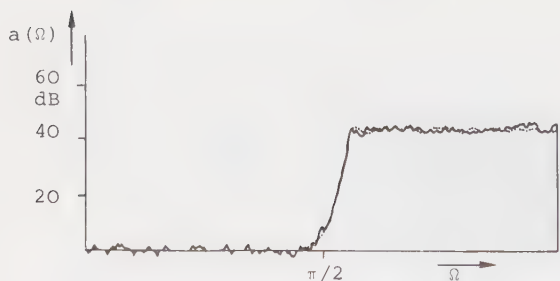


Bild 8.12c

$r = 8$

Bild 8.12: Vergleich der Dämpfungsverläufe bei der modifizierten Form

———— Simulation

..... Berechnung nach linearem Ersatzmodell

Es zeigt sich in beiden Fällen $r = 8$ eine sehr gute Übereinstimmung; ebenso für $r = 2$ in der modifizierten Form. Bei den übrigen Fällen, $r = 2$ im Zweierkomplement und bei den Versionen $r = 3$, decken sich die Verläufe im Bereich von Dämpfungsminima recht genau, während Dämpfungsspitzen bei den berechneten Kurven in auffälliger Weise abgeschnitten werden.

Die Erklärung dafür liegt in den vereinfachenden Annahmen des linearen Ersatzmodells: Hier wird der Fehleranteil $\varepsilon(k)$ als mit den Variablen "unkorrelierte" Rauschgröße betrachtet, die bei dieser Vorstellung als spektral weiß anzunehmen ist. In Wirklichkeit enthält $\varepsilon(k)$ aber Korrelationen höherer Ordnung, die von dem einfachen Ersatzmodell nicht erfaßt werden. Um die Auswirkungen dieser Korrelationen auf die Spektralverteilung zu untersuchen, wird für das Beispiel $r = 2$ (Zweierkomplement) das Leistungsspektrum der Fehlergröße $\varepsilon(k) = \Delta y - \sum_{\rho=1}^R \alpha_{\rho}^* x_{\rho}$ durch Simulation berechnet (die Restkoeffizientenfehler α_{ρ}^* können dabei vorab nach Gl. 4.66 bestimmt werden).

Das Ergebnis in Bild 8.13b zeigt, daß $\varepsilon(k)$ durchaus nicht spektral weiß ist, sondern eine starke Welligkeit aufweist.

Die Erklärung der Abweichungen im Dämpfungsverlauf nach dem linearen Ersatzmodell geht unmittelbar aus dem Vergleich der Bilder 8.13a und 8.13b hervor: In den Bereichen $\Omega_1 < \Omega < \Omega_2$ und $\Omega_5 < \Omega < \pi$ liegen Dämpfungsmaxima von $\varepsilon(k)$ vor, die Dämpfungsspitzen in Bild 8.13a werden bei Einsetzen des konstanten Wertes nach dem linearen Modell $a_{\varepsilon}(\Omega) = 39.1 \text{ dB}$, abgeschnitten. Umgekehrt fällt der Dämpfungsverlauf von $\varepsilon(k)$ im Bereich $\Omega_3 < \Omega < \Omega_4$ unter den beim linearen Ersatzmodell angenommenen konstanten Wert. Dadurch liegt in Bild 8.13a die berechnete Kurve über dem durch Simulation ermittelten Verlauf.

Allgemein ist zur Brauchbarkeit des linearen Ersatzmodells zur analytischen Berechnung realer Dämpfungsgänge folgendes festzustellen:

Für kleine Werte r wirkt sich die Annahme von weißem Rauschen für die Größe $\varepsilon(k)$ als Fehler aus. Dabei werden bei den betrachteten Beispielen die Dämpfungsspitzen abgeschnitten, die Dämpfungsminima jedoch richtig beschrieben; das Modell ist also zur Berechnung etwa der minimalen Sperrdämpfung eines Filters geeignet. Bei Erhöhung von r nimmt der Anteil N_{ε} zu, wird aber gleichzeitig spektral zusehens gleichmäßiger

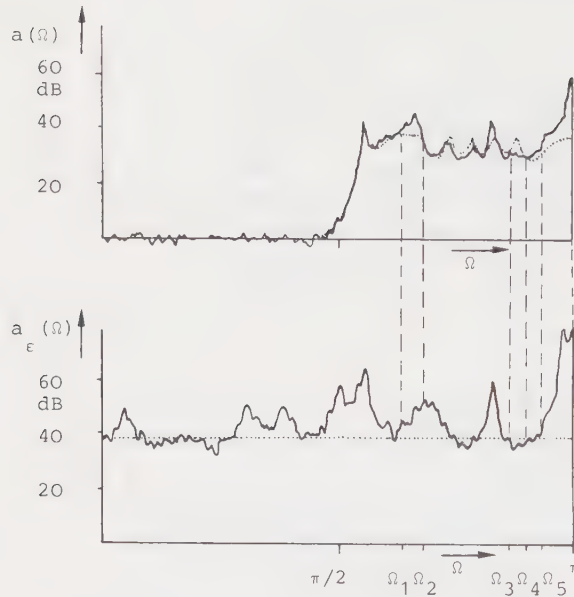


Bild 8.13a:

Filter-Dämpfungsgang

(Zweierkomplement, $r=2$)

Bild 8.13b

Spektralverteilung von $\epsilon(k)$:

$$a_{\epsilon}(\Omega) = -10 \log\{|E(\Omega)|^2 / \sigma_x^2\}$$

(mit $|E(\Omega)|^2$ = Leistungsspektrum von $\epsilon(k)$)

Bild 8.13: Erklärung der Abweichungen beim linearen Ersatzmodell

— Simulation
 lineares Ersatzmodell

verteilt. So wird z.B. im Fall $r = 8$ der reale Dämpfungsverlauf des Filters durch das Modell recht genau erfaßt. Eine Sonderstellung nimmt die modifizierte Form mit $r = 2$ ein: der Term $\epsilon(k)$ verschwindet hier, so daß das lineare Ersatzmodell das reale Filter exakt beschreibt.

8.4 Beurteilung der Realisierung nichtrekursiver Filter in Verteilter Arithmetik

Die Untersuchungen dieses Kapitels haben gezeigt, daß das Fehlerverhalten der Verteilten Arithmetik bei Anwendung auf nichtrekursive Filter offenbar günstig ist: So ließen sich für größere Werte r die Dämpfungsverzerrungen erheblich reduzieren; das Beurteilungsmaß der inneren Wortlängenerhöhung Δw bei der Betrachtung des gesamten Fehlerrauschens am Ausgang zeigte günstige Werte, insbesondere für die modifizierte Form.

Beim Vergleich ging man stets von der **Q u a n t i s i e -**
r u n g s s t u f e des Koeffizientenspeichers aus,
n i c h t v o n s e i n e r W o r t l ä n g e . So
wurde z.B. auch die "Wortlängenerhöhung" Δw nicht als Differenz von inneren und äußeren Wortlängen definiert, sondern als das logarithmierte Verhältnis der Quantisierungsstufen (vgl. Gl. 8.2).

Die Wortlänge des Koeffizientenspeichers bei der Verteilten Arithmetik ist jedoch im allgemeinen nicht wie im Fall der Konzentrierten Arithmetik durch

$$w_c = 2 + \lfloor \text{ld}(\text{Max}\{|a_p|\}) \rfloor - \text{ld } Q_c, \quad (8.10)$$

zu beschreiben (wobei das Symbol $\lceil a \rceil$ die größte ganze Zahl bedeutet, die kleiner oder gleich a ist), denn die Kombinationen von Koeffizienten können den Wertebereich der einzelnen Koeffizienten überschreiten und damit Übersteuerungsbits fordern. Allgemein ist die Koeffizientenspeicherwortlänge bei der Verteilten Arithmetik

im Fall der Zweierkomplement-Codierung

$$w_R = 2 + \left\lceil \text{ld} \left(\text{Max} \left\{ \sum_p a_p^{(+)}, \sum_p |a_p^{(-)}| \right\} \right) \right\rceil - \text{ld } Q_R \quad (8.11a)$$

mit $a_p^{(+)}$ = positive Koeffizienten, $a_p^{(-)}$ = negative Koeffizienten

und bei der modifizierten Form

$$w_R = 2 + \left[\lg \left(\sum_{\rho} |a_{\rho}| \right) \right] - \lg Q_R \quad (8.11b)$$

Im ungünstigsten Fall muß man hier mit $\lceil \lg r + 1 \rceil$ zusätzlichen Bits gegenüber der Konzentrierten Arithmetik rechnen, nämlich dann, wenn alle Koeffizientenwerte in der Nähe von $\max \{a_{\rho}\}$ liegen und r keine Zweierpotenz ist, z.B.:

$$r = 5, Q_R = Q_C = 2^{-7}$$

$$a_1 = a_2 = \dots = a_5 = 0.9$$

$$w_C = 2 + \lceil \lg 0.9 \rceil + 7 = 9 + \lceil -0.152 \rceil = 8 \text{ bit}$$

$$w_R = 2 + \lceil \lg 4.5 \rceil + 7 = 9 + \lceil 2.17 \rceil = 11 \text{ bit}$$

$$\lceil \lg r + 1 \rceil = \lceil 3.32 \rceil = 3 \text{ bit}$$

Bei der Anwendung auf nichtrekursive Filter ist die Anzahl der zusätzlichen Aussteuerungsbits abhängig von der Zusammenfassung der R/r Teil-Koeffizientengruppen. Dabei ist es in vielen Fällen günstig, die Werte der Impulsantwort unter Einhaltung möglichst großer zeitlicher Abstände zusammenzufassen, also entsprechend der in Kap. 3.2 vorgestellten Multiplex-Realisierung (vgl. z.B. die Pulsantwort des oben benutzten Tiefpasses, Bild 8.1a). Es sei daran erinnert, daß eine solche Gruppierung schon in einem anderen Zusammenhang von Interesse war: Bei der Fehleruntersuchung nichtrekursiver Filter, die von farbigem Rauschen erregt werden, war auf diese Weise die einfache analytische Behandlung nach Kap. 6 näherungsweise möglich (vgl. Kap. 7.1). Die Bedingungen für günstige Aussteuerung des Koeffizientenspeichers und die für eine einfache Fehlerbehandlung bei farbigen Eingangssignalen decken sich also.

Die Tabellen 8.1a und 8.1b zeigen bei der Quantisierungsstufe

$Q_R = 2^{-7}$ die erforderlichen Koeffizientenspeicher-Wortlängen für verschiedene r im Fall des oben untersuchten Tiefpaß-Beispiels. Zum Vergleich sind die zugehörigen minimalen Sperrdämpfungen und Gesamtleistungen der Ausgangsfehler infolge endlicher Koeffizientenspeicher-Wortlänge eingetragen.

r	Z w e i e r k o m p l .			m o d . F o r m		
	w_R	$\text{Min}\{a(\Omega)\}$ dB	$N_{\Delta Y}/Q_R^2$	w_R	$\text{Min}\{a(\Omega)\}$ dB	$N_{\Delta Y}/Q_R^2$
1	7	30.5	1.180	7	30.5	1.180
2	7	31.1	1.501	7	31.1	0.710
3	7	30.2	1.384	7	33.9	0.457
4	7	35.3	1.068	7	40.0	0.338
6	7	35.9	0.865	8	42.6	0.219
8	8	37.7	0.688	8	41.8	0.168

Tabelle 8.1: Wortlängen, minimale Sperrdämpfung, gesamte Fehlerleistung ($Q_R = 2^{-7}$) bei der Realisierung eines nichtrekursiven Tiefpasses 47. Grades

Nach dieser Aufstellung scheint die Realisierung in modifizierter Form mit $r = 4$ die günstigste Lösung zu sein, denn größere Werte r führen zwar zu geringeren Fehlern, erfordern dafür aber eine Wortlängenerhöhung des Koeffizientenspeichers um 1 Bit.

Diesen Überlegungen seien einige prinzipielle Bemerkungen hinzugefügt.

Aus obiger Tabelle ist nicht unmittelbar zu schließen, welche Realisierung bei einer bestimmten praktischen Anwendung zu wählen ist. Auf der Basis gleicher Speicher-Quantisierungsstufen wird zwar das Fehlerverhalten der Verteilten Arithmetik zunehmend günstiger, wobei jedoch hier als einziges Aufwandskriterium für eine Schaltungsrealisierung die Speicher-Wort-

länge berücksichtigt wird. Sinnvoll wäre es, den gesamten Aufwand unter Berücksichtigung des Fehlerverhaltens als Vergleichsmaß zu benutzen, indem z.B. an alle betrachteten Realisierungsformen die gleichen Anforderungen gestellt werden wie Erfüllung eines bestimmten Dämpfungstoleranzschemas, Begrenzung der Ausgangsrauschleistung oder auch Einhaltung von Geschwindigkeits-Bedingungen.

Derartige Fragen wurden in [8] eingehend behandelt, indem Kostenabschätzungen bei der Realisierung nichtrekursiver Filter in Verteilter Arithmetik vorhandenen Aufwandsmodellen für nichtrekursive Filter in Konzentrierter Arithmetik [34,39] gegenübergestellt wurden. Es zeigte sich, daß die Verteilte Arithmetik in dieser Hinsicht eine attraktive Realisierungsform darstellt.

Im vorliegenden Kapitel ging es also nicht so sehr darum, zu entscheiden, wie "gut" oder "schlecht" die eine oder andere Realisierungsform ist, etwa durch Vergleich der für eine bestimmte Fehlertoleranz erforderlichen Wortlängen — Gesamt-speicherbedarf bei Konzentrierter und Verteilter Arithmetik sind ohnehin nicht vergleichbar. Vielmehr interessierte das prinzipielle Fehlerverhalten bei der Realisierung nichtrekursiver Filter in Verteilter Arithmetik, und darin zeigten sich deutliche Unterschiede zur konventionellen Form der Konzentrierten Arithmetik wie die Einebnung des Dämpfungsverlaufs oder die Verschiebung von linearen Verzerrungen zu rauschartigen Fehlern.

Welche grundsätzlichen Fehlertypen dann bei konkreten Anwendungsfällen toleriert werden können, welche zu vermeiden sind, muß von Fall zu Fall individuell entschieden werden.

9. ZUSAMMENFASSUNG

In der vorliegenden Arbeit wird eine Analyse des Fehlers infolge endlicher Koeffizientenspeicher-Wortlänge bei der Verteilten Arithmetik vorgestellt. Dabei interessiert neben den Angaben der gesamten Fehlerleistung insbesondere die Zerlegung dieses Fehlers in Terme, die linear mit den Variablen korreliert sind - sie werden als Auswirkungen von Rest-Koeffizientenfehlern gedeutet - und einen unkorrelierten, rauschartigen Anteil.

Bei den Untersuchungen werden drei Klassen von Variablen unterschieden:

- weiße, gleichverteilte Rauschgrößen
- unkorrelierte Rauschgrößen mit beliebigen Spektral- und Amplitudenverteilungen
- beliebig untereinander korrelierte Signale

Für weißes, gleichverteiltes Rauschen gelingt eine geschlossene mathematische Formulierung von Gesamtfehlerleistung, Rest-Koeffizientenfehlern und der Leistung des unkorrelierten Rauschfehlers (Kap. 4). Unter Verwendung der abgeleiteten Formeln wird eine Untersuchung großer Ensembles von jeweils zufällig gewählten Koeffizientensätzen ausgeführt (Kap. 5); sie liefert allgemeine Aussagen über Erwartungswerte, Varianzen und Verteilungen der verschiedenen Fehleranteile sowie der Rest-Koeffizientenfehler. Es zeigt sich, daß die linear korrelierten Fehleranteile, also die Rest-Koeffizientenfehler mit anwachsender Länge der Koeffizientenvektoren zurückgehen, und zwar bei der modifizierten Form der Verteilten Arithmetik in stärkerem Maße als bei gewöhnlicher Zweierkomplement-Codierung der Variablen. Gleichzeitig läuft der Erwartungswert der Fehlerleistung bei Zweierkomplement-Codierung auf einen konstanten Endwert ein, der durch ein früher veröffentlichtes sehr einfaches Fehlermodell gegeben ist. Im Fall der modifizierten Form ist der Erwartungswert der Gesamtfehlerleistung für alle Vektorlängen konstant.

Bei Annahme von farbigen, beliebig verteilten Variablen ist eine analytische Fehlerbeschreibung dann möglich, wenn die Variablen untereinander unkorreliert sind (Kap. 6). Es ist dazu die Berechnung einer Bit-Korrelations-Matrix erforderlich, die die Korrelationen zwischen den Bits jeweils einer Variablen beschreibt. Sie läßt sich geschlossen aus der Verteilung der Variablen bestimmen. Es zeigt sich, daß die Rest-Koeffizientenfehler unverändert gegenüber denen bei weißem, gleichverteiltem Rauschen bleiben, während die Fehlerleistung von den spezifischen Eigenschaften der Variablen abhängt. Bei Ensemble-Untersuchungen ergibt sich für das hier verwendete Beispiel von Variablen-Färbung eine merkliche Reduktion der zu erwartenden Fehlerleistung gegenüber der bei Erregung mit weißem, gleichverteiltem Rauschen.

Für Variablen, die untereinander korreliert sind, gelingt eine befriedigende, geschlossene Formulierung des Fehlerverhaltens nicht mehr. Einen Vorteil allerdings zieht man aus der hier benutzten Beschreibung des Ausgangsfehlers, die nach dem Einfluß der Variablen und dem der Koeffizienten unterscheidet und so die Untersuchung großer Ensembles ermöglicht (Kap. 5). Man erkennt, daß die Rest-Koeffizientenfehler durch die Variablen-Korrelationen beeinflusst werden. Ihre Varianzen vergrößern sich bei allen betrachteten Beispielen verschiedener Variablen-Statistiken; ihr Wertebereich kann den der a-priori-Koeffizientenfehler überschreiten. Für die Erwartungswerte der Gesamtfehlerleistung führt bei den hier betrachteten Fällen die Vernachlässigung der Variablen-Korrelationen stets zu einer oberen Abschätzung.

Abschließend wird die für weiße, gleichverteilte Rausch-Erregung entwickelte Fehleranalyse auf die Untersuchung eines nichtrekursiven Filters angewendet (Kap. 8). Die Betrachtung der gesamten Ausgangs-Fehlerleistung führt zu allgemeinen Angaben über die notwendige Reduzierung der Koeffizientenspeicher-Quantisierungsstufe gegenüber der von Ein- und Ausgangssignalen. Zur Berechnung der linearen Verzerrungen des Filters werden die realen Dämpfungsverläufe auf der Basis eines linearen Ersatzmodells bestimmt. Der Vergleich mit Simulationsergebnissen zeigt, daß dieses Modell hinreichend

genaue Aussagen über die Dämpfungsverzerrungen liefert.

Wie in den vorangegangenen Abschnitten gilt auch für das Fehlerverhalten von nichtrekursiven Filtern in Verteilter Arithmetik generell, daß mit der Vergrößerung der Teilvektor-Längen die linearen Verzerrungen zurückgehen zugunsten eines rauschartigen Fehlers. Dadurch wird der Verlauf des realen Dämpfungsgangs eingeebnet, was gegenüber der Realisierung in Konzentrierter Arithmetik bei gleicher Koeffizientenspeicher-Quantisierungsstufe zu einer Verbesserung der minimalen Sperrdämpfung führt. Die modifizierte Form zeigt auch hier im Vergleich zur gewöhnlichen Zweierkomplement-Codierung in allen Fällen wieder günstigere Fehlereigenschaften.

A N H A N G 1

Zur Berechnung des Fehlers Δy_0

Mit Einsetzen von Gl. (4.31) in (4.10a) folgt:

$$\begin{aligned} \Delta y_0 = & - \sum_{\rho=1}^r \alpha_{\rho} \xi_{\rho 0} - Q_R \sum_{n=0}^{2^r-1} (e_n^{(0)} \cdot \prod_{\rho=1}^r (\xi_{\rho 0} \ominus v_{\rho})) \\ & + \sum_{i=1}^{w-1} 2^{-i} \left\{ \sum_{\rho=1}^r \alpha_{\rho} \xi_{\rho i} + Q_R \sum_{n=0}^{2^r-1} (e_n^{(0)} \prod_{\rho=1}^r (\xi_{\rho i} \ominus v_{\rho})) \right\}. \quad (A1.1) \end{aligned}$$

Nach Umordnung erhält man

$$\begin{aligned} \Delta y_0 = & \sum_{\rho=1}^r \alpha_{\rho} (-\xi_{\rho 0} + \sum_{i=1}^{w-1} 2^{-i} \xi_{\rho i}) + \\ & + Q_R \sum_{n=0}^{2^r-1} \left\{ e_n^{(0)} \left(- \prod_{\rho=1}^r (\xi_{\rho 0} \ominus v_{\rho}) + \sum_{i=1}^{w-1} 2^{-i} \prod_{\rho=1}^r (\xi_{\rho i} \ominus v_{\rho}) \right) \right\} \end{aligned} \quad (A1.2)$$

Mit Gl. (2.3) und der Definition (A1.3)

$$B(n, \mathbf{x}) = - \prod_{\rho=1}^r (\xi_{\rho 0} \ominus v_{\rho}) + \sum_{i=1}^{w-1} 2^{-i} \prod_{\rho=1}^r (\xi_{\rho i} \ominus v_{\rho}) \quad (A1.3)$$

ergibt sich für den Fehler

$$\Delta y_0 = \sum_{\rho=1}^r \alpha_{\rho} x_{\rho} + Q_R \sum_{n=0}^{2^r-1} e_n^{(0)} \cdot B(n, \mathbf{x}) \quad (A1.4)$$

A N H A N G 2

Zur Berechnung des Fehlers Δy_1

Gl. (4.36) wird in Gl. (4.11a) eingesetzt:

$$\begin{aligned} \Delta y_1 = & -\frac{1}{2} \left[\sum_{\rho=1}^r (\xi_{\rho 0} - \bar{\xi}_{\rho 0}) \alpha_{\rho} - \left[\sum_{\rho=1}^r (\xi_{\rho 0} - \bar{\xi}_{\rho 0}) \alpha_{\rho} \right] Q_R \right] \\ & + \frac{1}{2} \sum_{i=1}^{w-1} 2^{-i} \left[\sum_{\rho=1}^r (\xi_{\rho i} - \bar{\xi}_{\rho i}) \alpha_{\rho} - \left[\sum_{\rho=1}^r (\xi_{\rho i} - \bar{\xi}_{\rho i}) \alpha_{\rho} \right] Q_R \right] \\ & - 2^{-w} \left(\sum_{\rho=1}^r \alpha_{\rho} - \left[\sum_{\rho=1}^r \alpha_{\rho} \right] Q_R \right) \end{aligned} \quad (A2.1)$$

Nach Umordnung der Summanden

$$\begin{aligned} \Delta y_1 = & \sum_{\rho=1}^r \alpha_{\rho} \left(-\frac{1}{2} (\xi_{\rho 0} - \bar{\xi}_{\rho 0}) + \frac{1}{2} \sum_{i=1}^{w-1} 2^{-i} (\xi_{\rho i} - \bar{\xi}_{\rho i}) - 2^{-w} \right) + \\ & + \frac{1}{2} \left[\sum_{\rho=1}^r (\xi_{\rho 0} - \bar{\xi}_{\rho 0}) \alpha_{\rho} \right] Q_R - \frac{1}{2} \sum_{i=1}^{w-1} 2^{-i} \left[\sum_{\rho=1}^r (\xi_{\rho i} - \bar{\xi}_{\rho i}) \alpha_{\rho} \right] Q_R + \\ & + 2^{-w} \left[\sum_{\rho=1}^r \alpha_{\rho} \right] Q_R \end{aligned} \quad (A2.2)$$

erhält man mit Gl. (2.9b)

$$\begin{aligned} \Delta y_1 = & \sum_{\rho=1}^r \alpha_{\rho} x_{\rho} + \frac{1}{2} \left[\sum_{\rho=1}^r (\xi_{\rho 0} - \bar{\xi}_{\rho 0}) \alpha_{\rho} \right] Q_R - \\ & - \frac{1}{2} \sum_{i=1}^{w-1} 2^{-i} \left[\sum_{\rho=1}^r (\xi_{\rho i} - \bar{\xi}_{\rho i}) \alpha_{\rho} \right] Q_R + 2^{-w} \left[\sum_{\rho=1}^r \alpha_{\rho} \right] Q_R \end{aligned} \quad (A2.3)$$

A N H A N G 3

Berechnung des Ausdrucks $\langle B(n, \mathbf{x}(k)) \cdot B(m, \mathbf{x}(k)) \rangle$

Mit Gl. (2.10) erhält man

$$\begin{aligned} & \langle B(n, \mathbf{x}(k)) \cdot B(m, \mathbf{x}(k)) \rangle = \\ & \left\langle \left[- \prod_{\rho=1}^r (\xi_{\rho 0}(k) \oplus v_{\rho}) + \sum_{i=1}^{w-1} (2^{-i} \prod_{\rho=1}^r (\xi_{\rho i}(k) \oplus v_{\rho})) \right] \cdot \right. \\ & \quad \left. \cdot \left[- \prod_{\rho=1}^r (\xi_{\rho 0}(k) \oplus \mu_{\rho}) + \sum_{i=1}^{w-1} (2^{-i} \prod_{\rho=1}^r (\xi_{\rho i}(k) \oplus \mu_{\rho})) \right] \right\rangle, \end{aligned} \quad (A3.1)$$

wobei v_{ρ} und μ_{ρ} die Bits der Binärdarstellung von n und m bedeuten.

Nach Ausmultiplizieren ergeben sich 4 Terme, die im folgenden getrennt voneinander berechnet werden.

1. Term

$$\begin{aligned} & \left\langle \prod_{\rho=1}^r (\xi_{\rho 0}(k) \oplus v_{\rho}) \cdot \prod_{\rho=1}^r (\xi_{\rho 0}(k) \oplus \mu_{\rho}) \right\rangle = \\ & = \prod_{\rho=1}^r \langle (\xi_{\rho 0}(k) \oplus v_{\rho}) \cdot (\xi_{\rho 0}(k) \oplus \mu_{\rho}) \rangle = \\ & = \begin{cases} \prod_{\rho=1}^r \langle \xi_{\rho 0}(k) \oplus v_{\rho} \rangle & \text{für } n = m \\ 0 & \text{für } n \neq m \end{cases} \end{aligned} \quad (A3.2)$$

Mit der Einführung des Kronecker-Symbols

$$\delta_{nm} = \begin{cases} 1 & \text{für } n = m \\ 0 & \text{für } n \neq m \end{cases}$$

und der Bedingung $\langle \xi_{\rho 0}(k) \oplus v_{\rho} \rangle = \frac{1}{2}$ nach Gl. (4.43a) erhält man aus (A3.2):

$$\prod_{\rho=1}^r \langle (\xi_{\rho 0}(k) \ominus v_{\rho}) \cdot (\xi_{\rho 0}(k) \ominus v_{\rho}) \rangle = \delta_{nm} \cdot 2^{-r} \quad (A3.3)$$

2. Term

Mittelungs- und Produkt-Symbole können vertauscht werden, da die Variablen $\xi_{\rho i}$ für verschiedene ρ und i als statistisch unabhängig vorausgesetzt werden:

$$\begin{aligned} & - \prod_{\rho=1}^r \langle (\xi_{\rho 0}(k) \ominus v_{\rho}) \cdot \sum_{i=1}^{w-1} (2^{-i} \prod_{\rho=1}^r (\xi_{\rho i}(k) \ominus v_{\rho})) \rangle = \\ & = - \prod_{\rho=1}^r \langle \xi_{\rho 0}(k) \ominus v_{\rho} \cdot \sum_{i=1}^{w-1} (2^{-i} \prod_{\rho=1}^r \langle \xi_{\rho i}(k) \ominus v_{\rho} \rangle) \rangle \end{aligned} \quad (A3.4)$$

Für hinreichend große Wortlänge w kann

$$\sum_{i=1}^{w-1} 2^{-i} = (1 - 2^{-(w-1)}) \approx 1$$

gesetzt werden, so daß man aus (A3.4) erhält:

$$- \prod_{\rho=1}^r \left(\frac{1}{2}\right) \cdot \prod_{\rho=1}^r \left(\frac{1}{2}\right) \cdot \sum_{i=1}^{w-1} 2^{-i} \approx -2^{-2r} \quad (A3.5)$$

3. Term

Nach Vertauschung von v_{ρ} und μ_{ρ} erhält man die gleiche Form wie im 2. Term. Man kann also den Ausdruck (A3.5) übernehmen:

$$- \prod_{\rho=1}^r \langle (\xi_{\rho 0}(k) \ominus \mu_{\rho}) \cdot \sum_{i=1}^{w-1} (2^{-i} \prod_{\rho=1}^r (\xi_{\rho i}(k) \ominus v_{\rho})) \rangle \approx -2^{-2r} \quad (A3.6)$$

4. Term

$$\begin{aligned} & \left\langle \sum_{i=1}^{w-1} (2^{-i} \prod_{\rho=1}^r (\xi_{\rho i}(k) \ominus v_{\rho})) \right\rangle \cdot \left\langle \sum_{i=1}^{w-1} (2^{-i} \prod_{\rho=1}^r (\xi_{\rho i}(k) \ominus \mu_{\rho})) \right\rangle = \\ & = \sum_{i=1}^{w-1} \sum_{j=1}^{w-1} 2^{-i-j} \prod_{\rho=1}^r \langle (\xi_{\rho i}(k) \ominus v_{\rho}) \cdot (\xi_{\rho j}(k) \ominus \mu_{\rho}) \rangle \end{aligned} \quad (A3.7)$$

Hier müssen die beiden Fälle $i = j$ und $i \neq j$ unterschieden werden:

$$\begin{aligned}
 & \sum_{i=1}^{w-1} \sum_{\substack{j=1 \\ i \neq j}}^{w-1} 2^{-i-j} \prod_{\rho=1}^r \langle (\xi_{\rho i}(k) \ominus \nu_{\rho}) \cdot (\xi_{\rho j}(k) \ominus \mu_{\rho}) \rangle = \\
 &= \sum_{i=1}^{w-1} \sum_{\substack{j=1 \\ i \neq j}}^{w-1} 2^{-i-j} \prod_{\rho=1}^r \langle (\xi_{\rho i}(k) \ominus \nu_{\rho}) \cdot (\xi_{\rho j}(k) \ominus \mu_{\rho}) \rangle + \\
 &+ \sum_{i=1}^{w-1} 2^{-2i} \prod_{\rho=1}^r \langle (\xi_{\rho i}(k) \ominus \nu_{\rho}) (\xi_{\rho i}(k) \ominus \mu_{\rho}) \rangle \quad (A3.8)
 \end{aligned}$$

Mit Gl. (4.43b) und (4.43a) ergeben sich die Ausdrücke (A3.9) und (A3.10):

$$\langle (\xi_{\rho i}(k) \ominus \nu_{\rho}) (\xi_{\rho j}(k) \ominus \mu_{\rho}) \rangle = \frac{1}{4} \quad \text{für } i \neq j \quad (A3.9)$$

$$\langle (\xi_{\rho i}(k) \ominus \nu_{\rho}) (\xi_{\rho i}(k) \ominus \mu_{\rho}) \rangle = \begin{cases} \frac{1}{2} & \text{für } \nu_{\rho} = \mu_{\rho} \\ 0 & \text{für } \nu_{\rho} = \bar{\mu}_{\rho} \end{cases} \quad (A3.10)$$

Da für hinreichend große Wortlänge w gilt

$$\sum_{i=1}^{w-1} 2^{-2i} = \frac{1}{3}(1 - 2^{-2(w-1)}) \approx \frac{1}{3}$$

und

$$\sum_{i=1}^{w-1} \sum_{\substack{j=1 \\ i \neq j}}^{w-1} 2^{-i-j} = \sum_{i=1}^{w-1} \sum_{j=1}^{w-1} 2^{-i-j} - \sum_{i=1}^{w-1} 2^{-2i} \approx 1 - \frac{1}{3} = \frac{2}{3},$$

erhält man aus Gl. (A3.8) den Ausdruck

$$\begin{aligned}
 & \left(\frac{1}{4}\right)^r \sum_{i=1}^{w-1} \sum_{\substack{j=1 \\ i \neq j}}^{w-1} 2^{-i-j} + \left(\frac{1}{2}\right)^r \cdot \delta_{nm} \sum_{i=1}^{w-1} 2^{-2i} \approx \\
 & \approx \frac{2}{3} 2^{-2r} + \frac{1}{3} 2^{-r} \delta_{nm} \quad (A3.11)
 \end{aligned}$$

Die Gleichungen (A3.3, A3.5, A3.6 und A3.11) führen zu dem

gesuchten Ausdruck:

$$\langle B(n, \mathbf{x}(k)) \cdot B(m, \mathbf{x}(k)) \rangle \approx$$

$$\delta_{nm} 2^{-r} - 2 \cdot 2^{-2r} + \frac{2}{3} 2^{-2r} + \frac{1}{3} \delta_{nm} 2^{-r}$$

$$= \frac{4}{3} 2^{-r} (\delta_{nm} - 2^{-r}) \quad (\text{A3.12})$$

A N H A N G 4

Berechnung des Ausdrucks $\langle x_{\rho}(k) \cdot B(n, \mathbf{x}(k)) \rangle$

Es werden die Gleichungen (2.3) und (4.33) eingesetzt:

$$\begin{aligned} \langle x_{\rho}(k) \cdot B(n, \mathbf{x}(k)) \rangle &= \\ &= \left\langle -\xi_{\rho 0}(k) + \sum_{i=1}^{w-1} \xi_{\rho i}(k) \cdot 2^{-i} \right\rangle \left\langle -\prod_{\sigma=1}^r (\xi_{\sigma 0}(k) \oplus v_{\sigma}) + \right. \\ &\quad \left. + \sum_{i=1}^{w-1} 2^{-i} \prod_{\sigma=1}^r (\xi_{\sigma i}(k) \oplus v_{\sigma}) \right\rangle \end{aligned} \quad (A4.1)$$

Die nach Ausmultiplizieren entstehenden 4 Terme werden getrennt berechnet.

1. Term

$$\begin{aligned} \langle \xi_{\rho 0}(k) \cdot \prod_{\sigma=1}^r (\xi_{\sigma 0}(k) \oplus v_{\sigma}) \rangle &= \\ &= \langle \xi_{\rho 0}(k) \cdot (\xi_{\rho 0}(k) \oplus v_{\rho}) \rangle \cdot \prod_{\substack{\sigma=1 \\ \sigma \neq \rho}}^r \langle \xi_{\sigma 0}(k) \oplus v_{\sigma} \rangle = \begin{cases} 2^{-r} & \text{für } v_{\rho}=1 \\ 0 & \text{für } v_{\rho}=0 \end{cases} \end{aligned} \quad (A4.2)$$

2. Term

$$\begin{aligned} \langle -\xi_{\rho 0}(k) \cdot \sum_{i=1}^{w-1} 2^{-i} \prod_{\sigma=1}^r (\xi_{\sigma i}(k) \oplus v_{\sigma}) \rangle &= \\ &= - \sum_{i=1}^{w-1} 2^{-i} \cdot \langle \xi_{\rho 0}(k) \rangle \cdot \prod_{\sigma=1}^r \langle \xi_{\sigma i}(k) \oplus v_{\sigma} \rangle = \\ &= -2^{-r-1} \sum_{i=1}^{w-1} 2^{-i} = -2^{-(r+1)} \end{aligned} \quad (A4.3)$$

unter Voraussetzung von hinreichend großer Wortlänge w.

3. Term

$$\begin{aligned}
 & - \left\langle \sum_{i=1}^{w-1} \xi_{\rho i}(k) \cdot 2^{-i} \cdot \prod_{\sigma=1}^r (\xi_{\sigma 0}(k) \oplus v_{\sigma}) \right\rangle = \\
 & = \sum_{i=1}^{w-1} 2^{-i} \cdot \langle \xi_{\rho i}(k) \rangle \cdot \prod_{\sigma=1}^r \langle \xi_{\sigma 0}(k) \oplus v_{\sigma} \rangle = 2^{-(r+1)} \quad (A4.4)
 \end{aligned}$$

4. Term

Die Fälle $i = j$ und $i \neq j$ müssen wieder getrennt betrachtet werden.

$$\begin{aligned}
 & \left\langle \sum_{i=1}^{w-1} \sum_{j=1}^{w-1} 2^{-i-j} \xi_{\rho i}(k) \cdot \prod_{\sigma=1}^r (\xi_{\sigma j}(k) \oplus v_{\sigma}) \right\rangle = \\
 & \sum_{i=1}^{w-1} \sum_{\substack{j=1 \\ i \neq j}}^{w-1} 2^{-i-j} \langle \xi_{\rho i}(k) \rangle \cdot \prod_{\sigma=1}^r \langle \xi_{\sigma j}(k) \oplus v_{\sigma} \rangle + \\
 & + \sum_{i=1}^{w-1} 2^{-2i} \langle \xi_{\rho i}(k) \rangle \cdot \prod_{\sigma=1}^r \langle \xi_{\sigma i}(k) \oplus v_{\sigma} \rangle \quad (A4.5)
 \end{aligned}$$

Unter Verwendung von (4.43a,b) erhält man (A4.6) und (A4.7)

$$\begin{aligned}
 & \left\langle \xi_{\rho i}(k) \cdot \prod_{\sigma=1}^r (\xi_{\sigma j}(k) \oplus v_{\sigma}) \right\rangle = \\
 & = \langle \xi_{\rho i}(k) \rangle \cdot \langle \xi_{\rho j}(k) \oplus v_{\rho} \rangle \cdot \prod_{\substack{\sigma=1 \\ \sigma \neq \rho}}^r \langle \xi_{\sigma j}(k) \oplus v_{\sigma} \rangle = \\
 & = 2^{-(r+1)} \quad \text{für} \quad i \neq j \quad (A4.6)
 \end{aligned}$$

$$\begin{aligned}
 & \left\langle \xi_{\rho i}(k) \cdot \prod_{\sigma=1}^r (\xi_{\sigma i}(k) \oplus v_{\sigma}) \right\rangle = \\
 & = \langle \xi_{\rho i}(k) \rangle \cdot \langle \xi_{\rho i}(k) \oplus v_{\rho} \rangle \cdot \prod_{\substack{\sigma=1 \\ \sigma \neq \rho}}^r \langle \xi_{\sigma i}(k) \oplus v_{\sigma} \rangle =
 \end{aligned}$$

$$= \begin{cases} 2^{-r} & \text{für } v_{\rho} = 1 \\ 0 & \text{für } v_{\rho} = 0 \end{cases} \quad (\text{A4.7})$$

(A4.6) und (A4.7) liefern den Ausdruck (A4.8) für den 4. Term.

$$\begin{aligned} & \left\langle \sum_{i=1}^{w-1} \sum_{j=1}^{w-1} 2^{-i-j} \xi_{\rho i}(k) \cdot \prod_{\sigma=1}^r (\xi_{\sigma j}(k) \oplus v_{\sigma}) \right\rangle \approx \\ & \approx \frac{2}{3} 2^{-(r+1)} + \frac{1}{3} 2^{-r} \cdot v_{\rho} = \\ & = \begin{cases} \frac{1}{3} 2^{-r+1} & \text{für } v_{\rho} = 1 \\ \frac{1}{3} 2^{-r} & \text{für } v_{\rho} = 0 \end{cases} \quad (\text{A4.8}) \end{aligned}$$

Mit (A4.2), (A4.3), (A4.4) und (A4.8) erhält man schließlich den gesuchten Gesamtausdruck.

$$\begin{aligned} \langle x_{\rho}(k) \cdot B(n, x(k)) \rangle &= 2^{-r} \cdot v_{\rho} \cdot 2^{-(r+1)} \cdot 2^{-(r+1)} + \begin{cases} \frac{1}{3} 2^{-r+1} & \text{für } v_{\rho} = 1 \\ \frac{1}{3} 2^{-r} & \text{für } v_{\rho} = 0 \end{cases} \\ &= \begin{cases} \frac{1}{3} \cdot 2^{-r+1} & \text{für } v_{\rho} = 1 \\ -\frac{1}{3} \cdot 2^{-r+1} & \text{für } v_{\rho} = 0 \end{cases} \quad (\text{A4.9}) \end{aligned}$$

A N H A N G 5

Berechnung der Leistung N_ϵ

$$N_\epsilon = \langle (\Delta y - \sum_{\rho=1}^r \alpha_\rho^* x_\rho)^2 \rangle =$$

$$\langle \Delta y^2 \rangle - 2 \sum_{\rho=1}^r \alpha_\rho^* \langle \Delta y \cdot x_\rho \rangle + \sum_{\rho=1}^r \sum_{\sigma=1}^r \alpha_\rho^* \alpha_\sigma^* \langle x_\rho \cdot x_\sigma \rangle \quad (A5.1)$$

Aus Gl. (4.64a) und (4.63a) folgt

$$\sum_{\rho=1}^r \alpha_\rho^* \langle \Delta y \cdot x_\rho \rangle = \sum_{\rho=1}^r \alpha_\rho^* \cdot \frac{\alpha_\rho^*}{3}$$

und
$$\sum_{\rho=1}^r \sum_{\sigma=1}^r \alpha_\rho^* \alpha_\sigma^* \langle x_\rho \cdot x_\sigma \rangle = \sum_{\rho=1}^r (\alpha_\rho^*)^2 \cdot \frac{1}{3},$$

und man erhält mit (A5.1):

$$N_\epsilon = \langle \Delta y^2 \rangle - \frac{1}{3} \sum_{\rho=1}^r (\alpha_\rho^*)^2 \quad (A5.2)$$

In (A5.2) werden (4.54) und (4.66) eingesetzt:

$$\begin{aligned} N_\epsilon &= \frac{1}{3} \sum_{\rho=1}^r \left(\alpha_\rho^2 - 2^{-r+2} Q_R \alpha_\rho \sum_{n=0}^{2^r-1} (-1)^{\left\lfloor \frac{2n}{2^\rho} \right\rfloor} e_n \right) + \\ &+ \frac{4}{3} Q_R^2 2^{-r} \left(\sum_{n=0}^{2^r-1} (e_n)^2 - 2^{-r} \left(\sum_{n=0}^{2^r-1} e_n \right)^2 \right) - \\ &- \frac{1}{3} \sum_{\rho=1}^r \left(\alpha_\rho^2 - Q_R 2^{-r+2} \alpha_\rho \sum_{n=0}^{2^r-1} (-1)^{\left\lfloor \frac{2n}{2^\rho} \right\rfloor} e_n + \right. \\ &+ \left. Q_R^2 2^{-2(r-1)} \left(\sum_{n=0}^{2^r-1} (-1)^{\left\lfloor \frac{2n}{2^\rho} \right\rfloor} e_n \right)^2 \right) \\ N_\epsilon &= \frac{4}{3} Q_R^2 2^{-r} \left(\sum_{n=0}^{2^r-1} (e_n)^2 - 2^{-r} \left(\sum_{n=0}^{2^r-1} e_n \right)^2 - \right. \\ &- \left. 2^{-r} \sum_{\rho=1}^r \left(\sum_{n=0}^{2^r-1} (-1)^{\left\lfloor \frac{2n}{2^\rho} \right\rfloor} e_n \right)^2 \right) \quad A5.3 \end{aligned}$$

A N H A N G 6

Berechnung der Verteilung von $\alpha_{n\ell}$ |30|

Zur Lösung des Problems werden zunächst nur positive Koeffizientenfehler betrachtet; die Erweiterung auf negative Werte ist aus Symmetriegründen unmittelbar möglich.

Problemstellung:

Gegeben sind zwei statistisch unabhängige Prozesse $\{\alpha_1\}$ und $\{\alpha_2\}$, die beide im Intervall $[0, b]$ gleichverteilt sein sollen:

$$p_{\alpha_1}(\beta) = p_{\alpha_2}(\beta) = \begin{cases} \frac{1}{b} & \text{für } 0 \leq \beta \leq b \\ 0 & \text{sonst} \end{cases} \quad (\text{A6.1})$$

Es werden nun aus beiden Prozessen willkürlich Paare von Elementen α_1 und α_2 herausgenommen und addiert. Dabei werden nur diejenigen berücksichtigt, deren Summe größer oder gleich b ist. Ihre Verteilung ist zu bestimmen.

Man formuliert zwei Aussagen A und B:

A : $\alpha_1 \leq \beta$, benötigt für die Verteilungsdichte-Funktion,

B : $\alpha_1 + \alpha_2 \geq b$, Nebenbedingung

und ermittelt zunächst die bedingte Wahrscheinlichkeit

$$P\{A/B\} = \frac{P\{AB\}}{P\{B\}} \quad (\text{A6.2})$$

Dabei bedeutet $P\{AB\}$ die Verbundwahrscheinlichkeit der Ereignisse A und B, die durch die Verteilungsdichtefunktionen $p_{\alpha_1}(\beta)$ und $p_{\alpha_2}(\beta)$ ausgedrückt werden soll.

$$P\{AB\} = P\{(\alpha_1 + \alpha_2 \geq b) \wedge (\alpha_1 \leq \beta)\} \quad (\text{A6.3})$$

Für ein festgehaltenes $\xi \leq \beta$ (Bedingung A) schreibt man für (A6.3)

$$P\{\xi + \alpha_2 \geq b\} = P\{\alpha_2 \geq b - \xi\} = \int_{b-\xi}^{\infty} p_{\alpha_2}(\eta) d\eta \quad (A6.4)$$

Daraus erhält man durch Integration über die Variable ξ einen Ausdruck für $P\{AB\}$. Bei statistischer Unabhängigkeit von α_1 und α_2 gilt:

$$P\{AB\} = \int_{-\infty}^b p_{\alpha_1}(\xi) \int_{b-\xi}^{\infty} p_{\alpha_2}(\eta) d\eta d\xi \quad (A6.5)$$

Die Lösung von Gl. (A6.5) kann für den hier betrachteten Spezialfall (vgl. A6.1) direkt hingeschrieben werden:

$$\int_{b-\xi}^{\infty} p_{\alpha_2}(\eta) d\eta = \begin{cases} \frac{1}{b} \cdot \xi & \text{für } 0 \leq \xi \leq b \\ 1 & \text{für } \xi \geq b \end{cases} \quad (A6.6)$$

und damit

$$P\{AB\} = \begin{cases} 0 & \text{für } \beta \leq 0 \\ \frac{1}{2} \frac{\beta^2}{b^2} & \text{für } 0 \leq \beta \leq b \\ \frac{1}{2} & \text{für } \beta \geq b \end{cases} \quad (A6.7)$$

Die Wahrscheinlichkeit für das Auftreten von B ist unmittelbar hinzuschreiben:

$$P\{B\} = P\{\alpha_1 + \alpha_2 > b\} = 1 - \int_{-\infty}^b p_{\alpha_1}(\xi) * p_{\alpha_2}(\xi) d\xi \quad (A6.8)$$

Die Faltung der beiden Gleichverteilungen liefert

$$p_{\alpha_1}(\beta) * p_{\alpha_2}(\beta) = \begin{cases} \frac{1}{b^2} \cdot \beta & \text{für } 0 \leq \beta \leq b \\ \frac{2}{b} - \frac{1}{b^2} \beta & \text{für } b \leq \beta \leq 2b \\ 0 & \text{sonst} \end{cases} \quad (A6.9)$$

so daß sich für (A6.8) ergibt

$$P\{B\} = 1 - \frac{1}{b^2} \int_0^b \beta d\beta = \frac{1}{2}. \quad (A6.10)$$

Aus (A6.7) und (A6.10) erhält man die gesuchte bedingte Wahrscheinlichkeit

$$P\{A/B\} = \begin{cases} 0 & \text{für } \beta \leq 0 \\ \frac{\beta^2}{b^2} & \text{für } 0 \leq \beta \leq b, \\ 1 & \text{für } \beta \geq b \end{cases} \quad (\text{A6.11})$$

woraus mit $p_{\alpha_1/B} = \frac{\partial P\{A/B\}}{\partial \beta}$ die Verteilungsdichte-Funktion der Koeffizienten $\alpha_{n\ell}$ folgt:

$$p_{\alpha_1/B} = p_{\alpha_{n\ell}} = \begin{cases} \frac{2}{b^2} \cdot \beta & \text{für } 0 \leq \beta \leq b \\ 0 & \text{sonst} \end{cases} \quad (\text{A6.12})$$

Die Erweiterung auf negative Koeffizientenfehler ist durch symmetrische Ergänzung von (A6.12) unmittelbar zu erreichen (vgl. Bild A6.1,2)

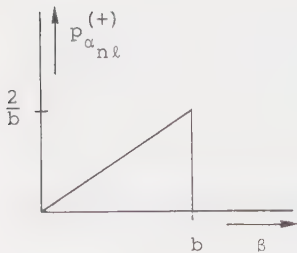


Bild A6.1:

Verteilung für positive
Koeffizientenfehler

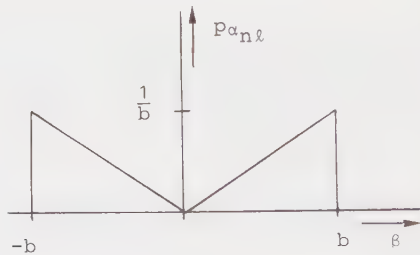


Bild A6.2:

Verteilung für positive und
negative Koeffizientenfehler

A N H A N G 7

Erwartungswert der Fehlerleistung bei der modifizierten Form

Mit Gl. (4.59) schreibt man für den Erwartungswert von $N_{\Delta y_1}$:

$$E\{N_{\Delta y_1}\} = \frac{1}{3} \sum_{\rho=1}^r \left\{ E\{\alpha_{\rho}^2\} - 2^{-r+2} Q_R \sum_{n=0}^{2^r-1} (-1)^{\left[\frac{2n}{2^{\rho}}\right]} E\{\alpha_{\rho} \cdot e_n^{(1)}\} \right\} + \\ + \frac{4}{3} Q_R^2 2^{-r} \sum_{n=0}^{2^r-1} E\{(e_n^{(1)})^2\} \quad (A7.1)$$

Wegen der Gleichverteilung der a-priori-Koeffizientenfehler gilt

$$E\{\alpha_{\rho}^2\} = Q_R^2/12 \quad (A7.2)$$

Zur Untersuchung des Ausdrucks $E\{\alpha_{\rho} \cdot e_n^{(1)}\}$ setze man die Definition des Fehlervektors (Gl. 4.38b) ein:

$$E\{\alpha_{\rho} e_n^{(1)}\} = \frac{1}{2Q_R} E\{\alpha_{\rho} [\pm \alpha_1 \pm \alpha_2 \dots \pm \alpha_{\rho} \dots \pm \alpha_r]_{Q_R}\} \quad (A7.3)$$

In dieser Gleichung kann selbstverständlich der Erwartungswert des Produktes nicht in ein Produkt von Erwartungswerten zerlegt werden, da beide Faktoren voneinander abhängen. Die Abhängigkeit liegt jedoch nur in einem Koeffizientenfehler α_{ρ} in der eckigen Klammer, so daß nur für diesen eine Vorzeichen-Fallunterscheidung gemacht wird. Zur Erläuterung wird das Beispiel $\alpha_{\rho} = \alpha_1$ betrachtet:

$$E\{\alpha_1 e_n^{(1)}\} = \frac{1}{2Q_R} E\{\alpha_1 [\pm \alpha_1 \pm \alpha_2 \dots \pm \alpha_r]_{Q_R}\} = \\ = \begin{cases} \frac{1}{2Q_R} \cdot E\{\alpha_1 [\pm \alpha_1 \pm \alpha_2 \dots \pm \alpha_r]_{Q_R}\} & \text{für } n \text{ gerade} \\ \frac{1}{2Q_R} \cdot E\{\alpha_1 [-\alpha_1 \pm \alpha_2 \dots \pm \alpha_r]_{Q_R}\} & \text{für } n \text{ ungerade} \end{cases} \quad (A7.4)$$

Wegen der Unabhängigkeit aller anderen Beiträge $\alpha_{\rho} (\rho \neq 1)$ von α_1

besitzt $E\{\alpha_1 e_n^{(1)}\}$ für alle geraden n den gleichen Wert, während sich für alle ungeraden n das Vorzeichen dieses Wertes umkehrt. Der Wert $e_n^{(1)}$ kann also durch eine beliebige Fehlervektor-Komponente (z.B. $e_o^{(1)}$) ersetzt werden:

$$E\{\alpha_1 e_n^{(1)}\} = \begin{cases} E\{\alpha_1 \cdot e_o^{(1)}\} & \text{für } n \text{ gerade} \\ -E\{\alpha_1 e_o^{(1)}\} & \text{für } n \text{ ungerade} \end{cases} \quad (\text{A7.5})$$

Allgemein erhält man für beliebige ρ

$$E\{\alpha_\rho \cdot e_n^{(1)}\} = (-1)^{\left\lfloor \frac{2n}{2^\rho} \right\rfloor} \cdot E\{\alpha_1 e_o^{(1)}\}, \quad (\text{A7.6})$$

also einen Ausdruck, der in der Vorzeichen-Formulierung an die Darstellung des Terms $\langle x_\rho \cdot B(n, \mathbf{x}) \rangle$ in Kapitel 4.4.3 (Gl. 4.53) erinnert.

Der dritte interessierende Term in Gl. (A7.1)

$$E\{(e_n^{(1)})^2\} = E\{([\pm\alpha_1 \pm \alpha_2 \pm \dots \pm \alpha_r]_{Q_R})^2\}$$

besitzt für alle n den gleichen Wert: Wegen der Gleichverteilung der α_ρ im Intervall $[-\frac{Q_R}{2}, +\frac{Q_R}{2}]$ sind positive und negative Vorzeichen gleich wahrscheinlich. Das allgemeine $e_n^{(1)}$ wird wieder durch $e_o^{(1)}$ ersetzt:

$$E\{(e_n^{(1)})^2\} = E\{(e_o^{(1)})^2\} \quad (\text{A7.7})$$

Mit (A7.2), (A7.6) und (A7.7) erhält Gl. (A7.1) die einfachere Form

$$\begin{aligned} E\{N_{\Delta y_1}\} &= \frac{r}{3} \left[\frac{Q_R^2}{12} - 2^{-r+2} Q_R (2^r \cdot E\{\alpha_1 e_o^{(1)}\}) \right] \\ &\quad + \frac{4}{3} Q_R^2 + 2^{-r} (2^r E\{(e_o^{(1)})^2\}) \\ &= \frac{Q_R^2}{36} \cdot r - \frac{4}{3} Q_R r E\{\alpha_1 e_o^{(1)}\} + \frac{4}{3} Q_R^2 E\{(e_o^{(1)})^2\} \end{aligned} \quad (\text{A7.8})$$

Empirisch wurden durch Ensemble-Mittelung folgende Ausdrücke gefunden:

$$E\{\alpha_1 e_0\} = \frac{Q_R}{24} \quad \forall \quad r \quad (A7.9)$$

$$E\{(e_0^{(1)})^2\} = \frac{r+1}{48} \quad (A7.10)$$

Nach Einsetzen von (A7.9, A7.10) in (A7.8) ergibt sich für $E\{N_{\Delta Y_1}\}$ ein für alle r konstanter Wert:

$$E\{N_{\Delta Y_1}\} = \frac{Q_R^2}{36} \cdot r - \frac{Q_R^2}{18} r + \frac{Q_R^2}{36} (r+1)$$

$$E\{N_{\Delta Y_1}\} = \frac{Q_R^2}{36} , \quad (A7.11)$$

womit das Ergebnis in Kap. 5.3.1 bestätigt ist.

A N H A N G 8

Berechnung von $\langle x_\rho(k) \cdot B(n, \mathbf{x}(k)) \rangle$ bei nicht gleich- verteilten Variablen

Nach Ausmultiplikation der Gl. (A4.1) ergeben sich 4 Terme:

$$\begin{aligned}
 \textcircled{1} &= \langle \xi_{\rho 0}(k) \cdot \prod_{\sigma=1}^r (\xi_{\sigma 0}(k) \oplus v_\sigma) \rangle \\
 \textcircled{2} &= - \sum_{i=1}^{w-1} 2^{-i} \langle \xi_{\rho 0}(k) \cdot \prod_{\sigma=1}^r (\xi_{\sigma i}(k) \oplus v_\sigma) \rangle \\
 \textcircled{3} &= - \sum_{i=1}^{w-1} 2^{-i} \langle \xi_{\rho i}(k) \cdot \prod_{\sigma=1}^r (\xi_{\sigma 0}(k) \oplus v_\sigma) \rangle \\
 \textcircled{4} &= \sum_{i=1}^{w-1} \sum_{j=1}^{w-1} 2^{-i-j} \langle \xi_{\rho i}(k) \cdot \prod_{\sigma=1}^r (\xi_{\sigma j}(k) \oplus v_\sigma) \rangle \quad (\text{A8.1})
 \end{aligned}$$

Für den ersten Term ist der Ausdruck (A4.2) unverändert zu setzen, da die Unabhängigkeit der Bits für verschiedene ρ nach wie vor gefordert wird und im übrigen keine Multiplikationen von verschiedenwertigen Bits auftreten:

$$\textcircled{1} = \begin{cases} 2^{-r} & \text{für } v_\rho = 1 \\ 0 & \text{für } v_\rho = 0 \end{cases} \quad (\text{A8.2})$$

Im zweiten Term sind die Korrelationen zwischen den Bits 0 und i (i $\in$ {1, ..., (w-1)}) zu berücksichtigen:

$$\langle \xi_{\rho 0}(k) \xi_{\rho i}(k) \rangle = \psi_{0i} \quad (\text{A8.3a})$$

$$\langle \xi_{\rho 0}(k) \cdot \bar{\xi}_{\rho i}(k) \rangle = \langle \xi_{\rho 0}(k) \cdot (1 - \xi_{\rho i}(k)) \rangle =$$

$$= \langle \xi_{\rho 0}(k) \rangle - \langle \xi_{\rho 0}(k) \cdot \xi_{\rho i}(k) \rangle = \frac{1}{2} - \psi_{0i} \quad (\text{A8.3b})$$

Es ergibt sich damit

$$\textcircled{2} = \begin{cases} -2^{-r+1} \sum_{i=1}^{w-1} \phi_{oi} 2^{-i} & \text{für } v_p = 1 \\ 2^{-r+1} \sum_{i=1}^{w-1} \phi_{oi} 2^{-i} - 2^{-r} & \text{für } v_p = 0 \end{cases} \quad (\text{A8.4})$$

bei hinreichend großer Wortlänge w (es wird wie in vorangegangenen Abschnitten $\sum_{i=1}^{w-1} 2^{-i} = 1$ gesetzt).

Für den dritten Term erhält man nach Vertauschung der Indizes 0 und i den gleichen Ausdruck:

$$\textcircled{3} = \begin{cases} -2^{-r+1} \sum_{i=1}^{w-1} \phi_{oi} 2^{-i} & \text{für } v_p = 1 \\ +2^{-r+1} \sum_{i=1}^{w-1} \phi_{oi} 2^{-i} - 2^{-r} & \text{für } v_p = 0 \end{cases} \quad (\text{A8.5})$$

Auch in den vierten Term werden die Elemente der Matrix ϕ eingesetzt. Unter Verwendung von (A8.3) erhält man

$$\textcircled{4} = \begin{cases} 2^{-r+1} \sum_{i=1}^{w-1} \sum_{j=1}^{w-1} \phi_{ij} 2^{-i-j} & \text{für } v_p = 1 \\ -2^{-r+1} \sum_{i=1}^{w-1} \sum_{j=1}^{w-1} \phi_{ij} 2^{-i-j+2^{-r}} & \text{für } v_p = 0 \end{cases} \quad (\text{A8.6})$$

Die Zusammenfassung von (A8.2, A8.4, A8.5 und A8.6) liefert den gesuchten Ausdruck:

$$\begin{aligned} & \langle x_p(k) \cdot B(n, x(k)) \rangle = \\ & = \begin{cases} 2^{-r} - 2^{-r+2} \sum_{i=1}^{w-1} \phi_{oi} 2^{-i} + 2^{-r+1} \sum_{i=1}^{w-1} \sum_{j=1}^{w-1} \phi_{ij} 2^{-i-j} & \text{für } v_p = 1 \\ -2^{-r} + 2^{-r+2} \sum_{i=1}^{w-1} \phi_{oi} 2^{-i} - 2^{-r+1} \sum_{i=1}^{w-1} \sum_{j=1}^{w-1} \phi_{ij} 2^{-i-j} & \text{für } v_p = 0 \end{cases} = \end{aligned}$$

$$= -(-1)^{\left[\frac{2n}{2^p}\right]} 2^{-r+1} \left(\frac{1}{2} - 2 \cdot \sum_{i=1}^{w-1} \psi_{oi} \cdot 2^{-i} + \sum_{i=1}^{w-1} \sum_{j=1}^{w-1} \psi_{ij} 2^{-i-j} \right) \quad (\text{A8.7})$$

Bei näherer Betrachtung erweist sich der in großen Klammern stehende Ausdruck als die Leistung des Eingangssignals $x_f(k)$.

$$\begin{aligned} \sigma_{x_f}^2 &= \langle x_\rho^2(k) \rangle = \langle \xi_{\rho 0}(k) \xi_{\rho 0}(k) \rangle - 2 \cdot \sum_{i=1}^{w-1} 2^{-i} \cdot \langle \xi_{\rho 0}(k) \cdot \xi_{\rho i}(k) \rangle + \\ &\quad + \sum_{i=1}^{w-1} \sum_{j=1}^{w-1} 2^{-i-j} \langle \xi_{\rho i}(k) \cdot \xi_{\rho j}(k) \rangle = \\ &= \frac{1}{2} - 2 \cdot \sum_{i=1}^{w-1} \psi_{oi} 2^{-i} + \sum_{i=1}^{w-1} \sum_{j=1}^{w-1} \psi_{ij} 2^{-i-j}, \quad (\text{A8.8}) \end{aligned}$$

so daß Gl. (A8.7) die einfache Form erhält:

$$\langle x_\rho(k) \cdot B(n, x(k)) \rangle = -(-1)^{\left[\frac{2n}{2^p}\right]} 2^{-r+1} \cdot \sigma_{x_f}^2 \quad (\text{A8.9})$$

A N H A N G 9

Berechnung von $\langle B(n, \mathbf{x}(k)) \cdot B(m, \mathbf{x}(k)) \rangle$ bei nicht gleichver- teilten Variablen

Gleichung (A3.1) liefert vier Terme

$$\begin{aligned} \textcircled{1} &= \prod_{\rho=1}^r \langle (\xi_{\rho 0}(k) \ominus v_{\rho}) (\xi_{\rho 0}(k) \ominus \mu_{\rho}) \rangle \\ \textcircled{2} &= - \sum_{i=1}^{w-1} 2^{-i} \prod_{\rho=1}^r \langle (\xi_{\rho 0}(k) \ominus v_{\rho}) \cdot (\xi_{\rho i}(k) \ominus \mu_{\rho}) \rangle \\ \textcircled{3} &= - \sum_{i=1}^{w-1} 2^{-i} \prod_{\rho=1}^r \langle (\xi_{\rho 0}(k) \ominus \mu_{\rho}) \cdot (\xi_{\rho i}(k) \ominus v_{\rho}) \rangle \\ \textcircled{4} &= \sum_{i=1}^{w-1} \sum_{j=1}^{w-1} 2^{-i-j} \prod_{\rho=1}^r \langle (\xi_{\rho i}(k) \ominus v_{\rho}) \cdot (\xi_{\rho j}(k) \ominus \mu_{\rho}) \rangle, \end{aligned} \quad (\text{A9.1})$$

von denen der erste wiederum von der Analyse für weißes, gleichverteiltes Eingangsrauschen übernommen werden kann (vgl. A3.3):

$$\textcircled{1} = \delta_{nm} 2^{-r} \quad (\text{A9.2})$$

Im zweiten Term werden zunächst die einzelnen Faktoren betrachtet. Mit Gl. (A8.3) gilt:

$$\langle (\xi_{\rho 0}(k) \ominus v_{\rho}) \cdot (\xi_{\rho i}(k) \ominus \mu_{\rho}) \rangle = \begin{cases} \cdot_{oi} & \text{für } v_{\rho} = \mu_{\rho} \\ \frac{1}{2} - \cdot_{oi} & \text{für } v_{\rho} = \bar{\mu}_{\rho} \end{cases} \quad (\text{A9.3})$$

Die Fallunterscheidung in (A9.3) muß bei der Bildung des Produkts über den Index ρ beachtet werden. Man führt dazu eine Größe

$$d_{nm} = \prod_{\rho=1}^r v_{\rho} \oplus \mu_{\rho} \quad (\text{A9.4})$$

ein, die die Anzahl der verschiedenen gleichwertigen Bits von n und m angibt. (Das Symbol $\oplus$ bedeutet die Antivalenz-Verknüpfung der Binärvariablen v_ρ und μ_ρ). Man erhält damit für den zweiten Term von (A9.1)

$$\begin{aligned} (2) &= - \sum_{i=1}^{w-1} (2^{-i} \psi_{oi}^{r-d_{nm}} (\frac{1}{2} - \psi_{oi})^{d_{nm}}) = \\ &= - \sum_{i=1}^{w-1} (2^{-i} \psi_{oi}^r (\frac{1}{2\psi_{oi}} - 1)^{d_{nm}}) \end{aligned} \quad (A9.5)$$

und den gleichen Ausdruck für den Term (3):

$$(3) = - \sum_{i=1}^{w-1} (2^{-i} \psi_{oi}^r (\frac{1}{2\psi_{oi}} - 1)^{d_{nm}}) . \quad (A9.6)$$

Der vierte Term schließlich führt nach den gleichen Überlegungen zu der Form

$$(4) = \sum_{i=1}^{w-1} \sum_{j=1}^{w-1} (2^{-i-j} \psi_{ij}^r (\frac{1}{2\psi_{ij}} - 1)^{d_{nm}}) , \quad (A9.7)$$

und mit den Gleichungen (A9.2, A9.5, A9.6 und A9.7) ergibt sich der gesuchte Ausdruck

$$\begin{aligned} &\langle B(n, \mathbf{x}(k)) \cdot B(m, \mathbf{x}(k)) \rangle = \\ &= \delta_{nm} \cdot 2^{-r} - 2 \sum_{i=1}^{w-1} 2^{-i} \psi_{oi}^r (\frac{1}{2\psi_{oi}} - 1)^{d_{nm}} + \\ &+ \sum_{i=1}^{w-1} \sum_{j=1}^{w-1} 2^{-i-j} \psi_{ij}^r (\frac{1}{2\psi_{ij}} - 1)^{d_{nm}} \end{aligned} \quad (A9.8)$$

In Gl. (A9.8) treten bei einigen Summengliedern unbestimmte Ausdrücke der Form $(0)^0$ auf. Z.B. ist für $i = j$ wegen

$$\psi_{ii} = \frac{1}{2} \quad \text{im Fall } n = m$$

$$\psi_{ii}^r (\frac{1}{2\psi_{ii}} - 1)^{d_{nm}} = 2^{-r} (0)^0 \stackrel{!}{=} 2^{-r} .$$

Die Glieder $(\frac{1}{2\psi_{ij}} - 1)^0$ sind also in jedem Fall gleich 1 zu setzen, auch wenn der Klammerausdruck verschwindet.

Anhang 10

Zusammenstellung der benutzten Beispiele

Beispiel 1 ($r=2$)

$$a_1 = 0.2178328 \quad \alpha_1 = -0.1174 Q_R$$

$$a_2 = -0.3159031 \quad \alpha_1 = -0.4356 Q_R$$

Zweierkomplement

Variablen	weiß gleichvert.	farbig, uncorr.		
		x_{f1}	x_{f2}	x_{f3}
Kapitel	4.7.1	6.7	7.3.2	7.3.2
$N_{\Delta y_O} / Q_R^2$	0.13351	0.03964	0.04358	0.03398
N_{ϵ_O} / Q_R^2	0.08333	0.02905	0.03363	0.02507
$\alpha_1^{*(0)} / Q_R$	0.3826	0.3826	0.3826	0.3826
$\alpha_2^{*(0)} / Q_R$	0.0644	0.0644	0.0644	0.0644

Variablen	korreliert		
	x_{f1}	x_{f2}	x_{f3}
Kapitel	7.3.2	7.3.2	7.3.2
$N_{\Delta y_O} / Q_R^2$	0.03830	0.02865	0.02026
N_{ϵ_O} / Q_R^2	0.02640	0.01601	0.01402
$\alpha_1^{*(0)} / Q_R$	0.3812	0.3827	0.3855
$\alpha_2^{*(0)} / Q_R$	0.0638	0.0653	0.0644

Modifizierte_Form

Variablen	weiß gleichvert.	farbig x_{f1} unkorr.	korreliert x_{f1}
Kapitel	4.7.1	6.7	7.3.2
$N_{\Delta Y_1}/Q_R^2$	0.05018	0.01023	0.01174
N_{ε_1}/Q_R^2	0.0	0.0	0.0
$\alpha_1^{*(1)}/Q_R$	0.3826	0.3826	0.3827
$\alpha_2^{*(1)}/Q_R$	0.0644	0.0644	0.0644

Beispiel 2 (r=2)

$$a_1 = 0.4941047$$

$$\alpha_1 = 0.2454 Q_R$$

$$a_2 = 0.1852891$$

$$\alpha_2 = -0.2830 Q_R$$

Variablen	Zweierkomplement		Modifizierte Form	
	weiß gleichvert.	unkorreliert x_{f1}	weiß gleichvert.	unkorreliert x_{f1}
Kapitel	4.7.1	6.7	4.7.1	6.7
$N_{\Delta Y}/Q_R^2$	0.04677	0.00954	0.03730	0.00760
N_{ε}/Q_R^2	0.0	0.0	0.0	0.0
α_1^*/Q_R	0.2454	0.2454	-0.2546	-0.2546
α_2^*/Q_R	-0.2830	-0.2830	0.2170	0.2170

Beispiel 3 ($r=8$)

$a_1 = 0.251531$	$\alpha_1 = 0.196 Q_R$
$a_2 = 0.218546$	$\alpha_2 = -0.026 Q_R$
$a_3 = -0.405625$	$\alpha_3 = 0.080 Q_R$
$a_4 = 0.849391$	$\alpha_4 = -0.278 Q_R$
$a_5 = -0.260927$	$\alpha_5 = -0.399 Q_R$
$a_6 = 0.282398$	$\alpha_6 = 0.147 Q_R$
$a_7 = 0.081344$	$\alpha_7 = 0.412 Q_R$
$a_8 = 0.155047$	$\alpha_8 = 0.154 Q_R$

Variablen	Zweierkomplement		Modifizierte Form	
	weiß gleichvert.	unkorreliert x_{f1}	weiß gleichvert.	unkorreliert x_{f1}
Kapitel	4.7.1	6.7	4.7.1	6.7
N_{AY}/Q_R^2	0.10789	0.05168	0.028193	0.014940
N_E/Q_R^2	0.10610	0.05131	0.027850	0.014860
α_1^*/Q_R	0.0160	0.0160	-0.0232	-0.0232
α_2^*/Q_R	-0.0031	-0.0031	0.0047	0.0047
α_3^*/Q_R	-0.0211	-0.0211	-0.0132	-0.0132
α_4^*/Q_R	-0.0201	-0.0201	-0.0124	-0.0124
α_5^*/Q_R	-0.0410	-0.0410	-0.0019	-0.0019
α_6^*/Q_R	-0.0019	-0.0019	-0.0098	-0.0098
α_7^*/Q_R	0.0450	0.0450	0.0059	0.0059
α_8^*/Q_R	0.0060	0.0060	-0.0018	-0.0018

Beispiel 4 (r=4)

$$\begin{aligned} a_1 &= 0.24817188 & \alpha_1 &= -0.234 Q_R \\ a_2 &= -0.35846875 & \alpha_2 &= 0.116 Q_R \\ a_3 &= 0.17246875 & \alpha_3 &= 0.076 Q_R \\ a_4 &= 0.13389844 & \alpha_4 &= 0.139 Q_R \end{aligned}$$

Variablen: weiß, gleichverteilt

Kapitel: 4.7.2

Zweierkomplement

$N_{\Delta Y_O} / Q_R^2$	0.0311
N_{ϵ} / Q_R^2	0.0

$\alpha_1^{*(0)} / Q_R$	-0.234
$\alpha_2^{*(0)} / Q_R$	0.116
$\alpha_3^{*(0)} / Q_R$	0.076
$\alpha_4^{*(0)} / Q_R$	0.139

Beispiel 5 (r=6)

$$\begin{aligned} a_1 &= 0.52855938 & \alpha_1 &= -0.3444 Q_R \\ a_2 &= -0.20002969 & \alpha_2 &= 0.3962 Q_R \\ a_3 &= 0.65689141 & \alpha_3 &= 0.0821 Q_R \\ a_4 &= 0.35950313 & \alpha_4 &= -0.0164 Q_R \\ a_5 &= -0.12233516 & \alpha_5 &= 0.3411 Q_R \\ a_6 &= 0.41606094 & \alpha_6 &= 0.2558 Q_R \end{aligned}$$

Variablen: weiß, gleichverteilt

Kapitel: 4.7.2

Zweierkomplement

$N_{\Delta Y_O} / Q_R^2$	0.10688
N_{ϵ} / Q_R^2	0.09896

$\alpha_1^{*(0)} / Q_R$	0.0306
$\alpha_2^{*(0)} / Q_R$	0.0212
$\alpha_3^{*(0)} / Q_R$	0.0821
$\alpha_4^{*(0)} / Q_R$	-0.0164
$\alpha_5^{*(0)} / Q_R$	-0.0339
$\alpha_6^{*(0)} / Q_R$	-0.1192

Beispiel 6 (r=4)

$$\begin{aligned}
 a_1 &= 0.52855938 & \alpha_1 &= -0.3444 Q_R \\
 a_2 &= -0.20002969 & \alpha_2 &= 0.3962 Q_R \\
 a_3 &= -0.12233516 & \alpha_3 &= 0.3411 Q_R \\
 a_4 &= 0.41606094 & \alpha_4 &= 0.2558 Q_R
 \end{aligned}$$

Variablen: weiß, gleichverteilt

Kapitel: 4.7.2

Zweierkomplement

$N_{\Delta Y_O} / Q_R^2$	0.10454
N_{ϵ_O} / Q_R^2	0.09896

$\alpha_1^* (0) / Q_R$	0.0306
$\alpha_2^* (0) / Q_R$	0.0212
$\alpha_3^* (0) / Q_R$	-0.0339
$\alpha_4^* (0) / Q_R$	-0.1192

Beispiel 7 (r=4)

Kapitel: 7.3.2

$$\begin{aligned}
 a_1 &= 0.21606016 & \alpha_1 &= -0.3443 Q_R \\
 a_2 &= 0.13127578 & \alpha_2 &= -0.1967 Q_R \\
 a_3 &= 0.32799688 & \alpha_3 &= -0.0164 Q_R \\
 a_4 &= 0.12225313 & \alpha_4 &= -0.3516 Q_R
 \end{aligned}$$

Zweierkomplement

Modifizierte Form

Variablen	unkorreliert x_{f1}	korreliert x_{f1}	unkorreliert x_{f1}	korreliert x_{f1}
$N_{\Delta Y} / Q_R^2$	0.04669	0.04940	0.010603	0.004759
N_{ϵ} / Q_R^2	0.03728	0.02396	0.009351	0.003930
α_1^* / Q_R	0.1557	0.06809	0.03068	0.071406
α_2^* / Q_R	0.3032	0.61475	-0.07174	-0.165824
α_3^* / Q_R	-0.0164	-0.08866	0.10855	0.051052
α_4^* / Q_R	0.1484	0.02752	0.02338	0.040205

L I T E R A T U R

- [1] E. Andersson
A Digital Filter Implemented in Parallel Form
Symp. on Digital Filters Imperial College, London,
Aug. 1971
- [2] A. Croisier, D.J. Esteban, M.E. Levilion, V. Riso
Digital Filter for PCM Encoded Signals
US Patent 3777 130, Dez. 1973
- [3] W.D. Little
An Algorithm for High-Speed Digital Filters
IEEE Trans. on Comp., Bd. C-23, May 1974, S. 466-469
- [4] W.K. Giloi, K.D. Kammeyer
On the Design of All-Digital Components for Low-Speed and High-Speed Modems Using MOS Technology
IEEE International Conference on Communications,
Minneapolis, Juni 1974, S. 39C-1 bis 39C-5
- [5] A. Peled, B. Liu
A New Hardware Realization of Digital Filters
IEEE Trans. on ASSP, Bd. ASSP-22, Dez. 1974, S. 456-462
- [6] H.W. Schüßler
Einführungsvortrag zum DFG-Kolloquium
"Digitale Systeme zur Signalverarbeitung"
Erlangen, März 1974
- [7] M. Büttner, H.W. Schüßler
On Structures for the Implementation of the
Distributed Arithmetic
NTZ, Bd. 29, Juni 1976, S. 472 - 477
- [8] K.D. Kammeyer
Digital Filter Realization in Distributed Arithmetic
European Conf. on Circuit Theory and Design,
Genua, Sept. 1976, S. 717 - 725

- 9 | T.R.C.M. Claasen, W.F.G. Mecklenbräuker, J.D.H. Peek
Some Considerations on the Implementation of Digital
Systems for Signal Processing
Philips Res. Rept. 30, 1975, S. 73 - 84
- | 10 | B. Liu, A. Peled
A New Hardware Realization of High-Speed Fast
Fourier Transformers
IEEE Trans. on ASSP, Bd. ASSP-23, Nr. 6, 1975,
S. 543 - 547
- | 11 | L.H. Goldstein, B. Liu
An ADPCM Realization of Nonrecursive Digital Filters
IEEE Trans. on ASSP, Bd. ASSP-24, Nr. 4, 1976,
S. 312 - 320
- | 12 | B. Liu, A. Peled
Digital Signal Processing
John Wiley & Sons, New York 1976
- | 13 | S. White
On the Mechanization of Vector Multiplication
Proc. IEEE, Bd. 63, 1975, S. 730 - 731
- | 14 | K.D. Kammeyer
Realisierung nichtrekursiver Filter in Verteilter
Arithmetik
DFG-Kolloquium "Digitalfilter", Bochum, Okt. 1975
- | 15 | H.W. Schüßler
Ein Gütemaß für die Realisierung digitaler Filter
DFG-Kolloquium "Digitalfilter", Bochum Okt. 1975
- | 16 | M. Büttner
Some Improvements in Digital Hardwired Approximation
of the Amplitude of Quadrature Pairs
AEÜ, Bd. 30, 1970, S. 167 - 171

- [17] K.D. Kammeyer
Realisierung digitaler Datensender mit programmier-
barer Modulationsform
DFG-Kolloquium "Digitale Übertragung", Stuttgart,
Feb. 1977

- [18] A. Peled, B. Liu, K. Steiglitz
A Note on Implementation of Digital Filters
IEEE Trans. on ASSP, Bd. ASSP-23, 1975, S. 387 - 388

- [19] C.S. Burrus
Digital Filter Realization by Distributed Arithmetic
International Symposium on Circuits and Systems
München, April 1976, S. 106 - 109

- [20] S. Zohar
New Hardware Realizations of Nonrecursive Digital
Filters
IEEE Trans. on Computers, Bd. C-22, Nr. 4, 1973,
S. 328 - 338

- [21] S. Zohar
The Counting Recursive Digital Filter
IEEE Trans. on Computers, Bd. C-22, Nr. 4, 1973,
S. 338 - 347

- [22] Trần-Thống
Finite Wordlength Effects in the ROM Digital Filter
IEEE Trans. on ASSP, Bd. ASSP-24, 1976, S. 436 - 437

- [23] K.D. Kammeyer
Hardware-Realisierungen digitaler Systeme in Daten-
übertragungsgeräten
DFG-Kolloquium "Digitale Systeme zur Signalverarbeitung"
Erlangen, März 1974

- [24] H.W. Schüßler
Persönliche Mitteilungen 1976

- |25| K.D. Kammeyer
Quantization Error Analysis of the Distributed
Arithmetic
Zur Publikation eingereicht bei IEEE Trans. on Circ.
and Syst.
- |26| R.C. Booton
Nonlinear Control Systems with Random Inputs
IRE-Trans. on Circuit Theory, 1954, S. 9 - 18
- |27| H. Schlitt
Stochastische Vorgänge in linearen und nichtlinearen
Regelkreisen
Braunschweig, Vieweg, Verlag, 1968
- |28| B. Eckhardt
Untersuchungen des Multipliziererrohrs in
digitalen Filtern
Ausgew. Arbeiten über Nachrichtensysteme, Hrsg. von
W. Schüßler, Univ. Erlangen-Nürnberg, Nr. 21, 1975
- |29| B. Eckhardt, H.W. Schüßler
On the Quantization Error of a Multiplier
Int. Symp. on Circuits and Systems,
München, 1976, S. 634 - 637
- |30| G. Steeger
Persönliche Mitteilung 1976
- |31| H.W. Schüßler
Persönliche Mitteilung 1977
- |32| D.S.K. Chan, L.R. Rabiner
Analysis of Quantization Errors in the Direct Form
for Finite Impulse Response Digital Filters
IEEE Trans. on Audio Acc. Bd. AU-21, 1973, S. 354 - 366

- |33| U. Heute
Necessary and Efficient Expenditure for Nonrecursive
Digital Filters in Direct Structure
Europ. Conf. Circ. Th. and Design, London 1974,
IEE publ. No. 116, S. 13 - 19

- |34| U. Heute
Über Realisierungsprobleme bei nicht-rekursiven
Digitalfiltern
Ausgew. Arbeiten über Nachrichtensysteme, Hrsg. von
W. Schüßler, Univ. Erlangen-Nürnberg, Nr. 20, 1975

- |35| T.W. Parks, J.H. McClellan
A Program for the Design of Linear Phase Finite
Impulse Response Digital Filters
IEEE Trans. on Audio Acc., Bd. AU-20, 1972, S. 195 -199

- |36| H.W. Schüßler
Digitale Systeme zur Signalverarbeitung
Springer-Verlag, Berlin 1973

- |37| H.D. Helms
Nonrecursive Digital Filters: Design Methods for
Achieving Specifications on Frequency Response
IEEE Trans. on Audio and Electroacoustics, Bd. AU-16
No. 3, 1968, S. 336 - 342

- |38| A. Papoulis
Minimum-Bias Windows for High-Resolution Spectral
Estimates
IEEE Trans. on Inf.Th., Bd. IT-19, No. 1, 1973, S.9-12

- |39| U. Heute
Best Realization of Flexible Digital FIR Filters
European Conf. on Circuit Theory and Design,
Genua, Sept. 1976, S. 741 - 747

L e b e n s l a u f

15.10.1944	geboren in Ahlhorn bei Oldenburg
1951 - 1957	Grundschule in Bremen-Grohn
1957 - 1964	Gerhard-Rohlfs-Gymnasium in Bremen- Veegesack (math. naturwissenschaftl. Zweig)
1964	Abitur
1964	Vorpraktikum (26 Wochen) zum Studium der Elektrotechnik (Bremer-Woll-Kämmerei, Bremen-Blumenthal)
1964 - 1971	Studium der Elektrotechnik an der Techn. Universität Berlin (Fachrichtung Nach- richtentechnik)
1967	Diplom-Vorprüfung
1968	Fachpraktikum (6 Wochen) bei der Firma Siemens, Berlin
1969	Fachpraktikum (8 Wochen) bei der Fa. AEG- Telefunken, Hannover
1971	Abschluß des Studiums mit Diplomarbeit (Informationsverarbeitung) und Diplom- Hauptprüfung an der TU Berlin
1972 - 1973	Wissenschaftlicher Mitarbeiter am Heinrich- Hertz-Institut, Berlin
1973 - 1974	Wissenschaftlicher Mitarbeiter, Informatik II an der Universität des Saarlandes, Saarbrücken
seit 1974	Wissenschaftlicher Mitarbeiter am Institut für Nachrichtentechnik an der Universität Erlangen- Nürnberg

Karl-Dirk Kammeyer

(Karl-Dirk Kammeyer)

Copy-Center 2000
852 ERLANGEN · Obere Karlstr. 30
Tel. 09131-26197